Quelques résultats d'équivalence asymptotique pour des expériences statistiques dans un cadre non paramétrique

Ester Mariucci

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathématiques Appliquées**

Arrêté ministériel : 7 août 2006

Présentée par

Ester Mariucci

Thèse dirigée par **Sana Louhichi**
et codirigée par **Pierre Étoré**

préparée au sein du **Laboratoire Jean Kuntzmann**
et de l'**École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique**

Quelques résultats d'équivalence asymptotique pour des expériences statistiques dans un cadre non paramétrique

Thèse soutenue publiquement le **16 septembre 2015**,
devant le jury composé de :

Mme Valentine Genon-Catalot
Professeur, Université Paris Descartes, Présidente
M. Ion Grama
Professeur, Université de Bretagne-Sud, Rapporteur
M. Marc Hoffmann
Professeur, Université Paris-Dauphine, Rapporteur
M. Anatoli Juditsky
Professeur, Université Joseph Fourier, Examinateur
Mme Sana Louhichi
Professeur, Université Joseph Fourier, Directeur de thèse
M. Pierre Étoré
Maître de conférences, Université Grenoble INP, Co-Directeur de thèse

À Marco et Francesco

Remerciements

Mes premiers remerciements s'adressent à mes directeurs de thèse, Sana Louhichi et Pierre Étoré. Je vous suis infiniment reconnaissante de m'avoir fait confiance il y a trois ans en acceptant de m'encadrer en thèse. Du stage de master jusqu'à la rédaction du manuscrit, vous m'avez guidée tout en me laissant l'autonomie et la liberté nécessaire à tout travail de recherche. Merci de m'avoir toujours soutenue et encouragée et de m'avoir confortée dans mon envie de poursuivre ma carrière dans la recherche.

Je remercie chaleureusement Ion Grama et Marc Hoffmann d'avoir accepté de lire en détail ce long texte et pour les précieux conseils qu'ils m'ont donnés pour l'améliorer. Je suis honorée qu'ils aient accepté de rapporter ma thèse.

Je remercie également sincèrement Valentine Genon-Catalot et Anatoli Juditsky d'avoir accepté de faire partie de mon jury. Je tiens à remercier Valentine pour toutes nos discussions sur les processus de diffusion en petite variance. Je remercie Anatoli pour ses conseils précieux et pour m'avoir mise en contact avec des équipes de recherche très réputées dans mon domaine.

Je remercie très sincèrement Markus Reiss pour son accueil à Berlin, sa gentillesse et sa disponibilité. Cette thèse n'aurait pas contenu certains de ses résultats sans nos discussions, toujours très intéressantes et enrichissantes.

Je remercie Didier Piau, responsable du Master 1 Mathématiques de l'Institut Fourier à l'époque où je voulais poursuivre mes études en France. Sa disponibilité et son accueil chaleureux m'ont beaucoup aidée et motivée dans la continuation de cette aventure grenobloise.

Cette thèse a été préparée au sein du LJK et je tiens à remercier toutes les personnes qui travaillent à la tour IRMA pour avoir créé une ambiance de travail agréable et conviviale. Plus particulièrement, je souhaite exprimer ma gratitude à Stéphane Labbé, référent exemplaire de l'École Doctorale pour les mathématiques appliquées, à Adeline Leclercq Samson pour ses conseils précieux et sa disponibilité et à Marianne Clausel et Stéphane Girard pour m'avoir incitée et aidée à participer à de nombreuses conférences internationales. Je remercie aussi tous les chercheurs que j'ai rencontrés lors de séminaires ou de conférences et qui m'ont conseillée ainsi que tous les doctorants et postdoctorants. Parmi ces derniers, je tiens à remercier Charline, Charlotte, Gwennaëlle, Jean-Marc, Sebastian, Pablo, Claire et Émilie. Je voudrais adresser un remerciement particulier à Céline que j'ai eu le plaisir de rencontrer à plusieurs occasions : je garde un très bon souvenir des échanges scientifiques que nous avons eus.

Un grand merci au “quatrième étage” de la tour et au service informatique pour leur amabilité et leur efficacité, notamment à Hélène, Laurence, Juana, Patrice et Frédéric.

De même, je voudrais dédier une pensée spéciale à toutes les personnes qui ont contribué à me former au métier d'enseignant-chercheur, autant les animateurs des formations du Label RES que les collègues avec lesquels j'ai partagé les cours ainsi que les étudiants. En particulier, merci à Hervé Pajot, mon responsable pédagogique, qui a toujours été disponible et à l'écoute lorsque j'en ai eu besoin.

Et me voilà arrivée à la partie la plus difficile ! Comment puis-je remercier avec une poignée de mots tous ceux avec lesquels j'ai partagé, au quotidien presque, ces trois dernières années ? J'aurais aimé vous remercier un par un, fixer sur le papier les souvenirs des moments passés ensemble, raconter les petites choses, gentillesses, blagues qui ont changé mes journées et vous dire à quel point tout cela a eu un effet bénéfique sur la réalisation de ce manuscrit. Je ne le ferai pas car il ne me semble pas judicieux de présenter un manuscrit où la section des remerciements est plus longue que celles contenant les résultats principaux de la thèse. Soyez compréhensifs et acceptez ces remerciements concis mais écrits avec le cœur.

Merci donc à tous les thésards, postdoc, ATER du LJK et de l'INRIA (et à ceux qui ne le sont plus) : Cécile, Meryam, Pierre-Olivier, Farida, Charlotte, Matthias, Chloé, Julien, Morgane, Romain, Federico P., Alessandro, Jonathan, Gildas, Christine, Kévin, Pierre-Jean, Federico Z., Souhil, Meriem, Margaux, Matteo, Thomas, Konstantina, Alexandre, Rémi, Charles, Nelson, Adrien, Nadia, Benoît, Roland, Achmad, Thi To Nhu, Pierre-Luc, Dmitri, Danila et Ratikanta.

Un grand merci à tous mes amis de Grenoble : Angelo, Bojo, Sara, Giorgio, Ilaria, Isabel, Marco, Marco, Federico, Filippo, Luca, Matthias, Francesco, Peter, Julia, Roberta, Tanja et Alex.

Merci aussi à la famille de mon mari Marco pour m'avoir accueillie si chaleureusement : je suis sincèrement touchée par votre amabilité et votre gentillesse et je suis très contente de pouvoir partager avec vous le jour de la soutenance.

Mes plus profonds remerciements s'adressent à toute ma famille et spécialement à ma maman Elena, à mon oncle Gianni et à ma grand-mère Dolores : je n'aurais jamais fait un doctorat en France sans leur soutien ! Merci de m'avoir toujours encouragée à m'investir dans les études et laissée libre de faire mes choix, même quand ils prévoyaient de m'éloigner significativement de chez nous. J'espère être toujours digne de la confiance que vous m'avez accordée. Une pensée particulière à mon grand-père Renato. Merci à ma cousine Michela de m'avoir écoutée et aidée à relâcher la pression, même à distance.

Enfin, merci à toi, Marco. Merci d'avoir été toujours là, dans les bons moments et les

mauvais. Pour m'avoir écoutée, soutenue, aidée et fait confiance au quotidien. Tes conseils ont été précieux, ta présence douce et ta patience juste incroyable. Mon cœur est plein d'amour et de sincère gratitude.

Résumé

Nous nous intéressons à l'équivalence asymptotique, au sens de Le Cam, entre différents modèles statistiques. Plus précisément, nous avons exploré le cas de modèles statistiques associés à l'observation discrète de processus à sauts ou de diffusions unidimensionnelles, ainsi que des modèles à densité plus classiques.

Ci-dessous, nous présentons brièvement les différents chapitres de la thèse.

Nous commençons par présenter tous nos résultats dans un premier chapitre introductif. Ensuite, dans le Chapitre 2 nous rappelons les points clés de la théorie de Le Cam sur les expériences statistiques en se plaçant dans un contexte non paramétrique.

Les Chapitres 3 et 4 traitent de l'équivalence asymptotique pour des modèles statistiques associés à l'observation discrète (haute fréquence) de processus à sauts. Dans un premier temps nous nous focalisons sur un problème d'équivalence en ce qui concerne l'estimation de la dérive, supposée appartenir à une certaine classe fonctionnelle. Il s'avère (Chapitre 3) qu'il y a une équivalence asymptotique, en ce qui concerne l'estimation de la dérive, entre le modèle statistique associé à l'observation discrète d'un processus additif X et le modèle statistique gaussien associé à l'observation discrète de la partie continue de X. Dans un deuxième temps, nous nous sommes intéressés au problème de l'estimation non paramétrique de la densité de Lévy f relative à un processus de Lévy à sauts purs, Y. Le Chapitre 4 illustre l'équivalence asymptotique, en ce qui concerne l'estimation de f, entre le modèle statistique associé à l'observation discrète de Y et un certain modèle de bruit blanc gaussien ayant $\sqrt{f}$ comme dérive.

Le Chapitre 5 présente une extension d'un résultat bien connu sur l'équivalence asymptotique entre un modèle à densité et un modèle de bruit blanc gaussien. Le Chapitre 6 étudie l'équivalence asymptotique entre un modèle de diffusion scalaire avec une dérive inconnue et un coefficient de diffusion qui tend vers zéro et le schéma d'Euler correspondant. Dans le Chapitre 7 nous présentons une majoration en distance L_1 entre les lois de processus additifs.

Le Chapitre 8 est consacré aux conclusions et discute des extensions possibles des travaux de thèse.

Abstract

The subject of this Ph.D. thesis is the asymptotic equivalence, in the Le Cam sense, between different statistical models. Specifically, we explore the case of statistical models associated with the discrete observation of jump processes or diffusion processes as well as more classical density models.

Below, we briefly introduce the different chapters of this dissertation.

We begin by presenting our results in a first introductory chapter. Then, in Chapter 2, we recall the key points of the Le Cam theory on statistical experiences focusing on a nonparametric context.

Chapters 3 and 4 deal with asymptotic equivalences for statistical models associated with discrete observation (high frequency) of jump processes. First, we focus on an equivalence problem regarding the estimation of the drift, assumed to belong to a certain functional class. It turns out (Chapter 3) that there is an asymptotic equivalence, for what concerns the estimation of the drift, between the statistical model associated with the discrete observation of an additive process X and the Gaussian statistical model associated with the discrete observation of the continuous part of X. Then we study the problem of nonparametric density estimation for the Lévy density f of a pure jump Lévy process Y. Chapter 4 illustrates the asymptotic equivalence, for what concerns the estimation of f, between the statistical model associated with the discrete observation of Y and a certain Gaussian white noise model having $\sqrt{f}$ as drift.

In Chapter 5 we present an extension of the well-known asymptotic equivalence between density estimation experiments and a Gaussian white noise model. Chapter 6 describes the asymptotic equivalence between a scalar diffusion model with unknown drift and with diffusion coefficient tending to zero and the corresponding Euler scheme. In Chapter 7 we present a bound for the L_1 distance between the laws of additive processes.

Chapter 8 is devoted to conclusions and discusses possible extensions of the results of this thesis.

Table des matières

Chapitre 1

Introduction

1.1 Équivalences asymptotiques au sens de Le Cam

1.1.1 Généralités

Une première notion en statistique mathématique est celle d'*expérience* (ou *modèle*) *statistique*. D'un point de vue formel, une *expérience statistique* est définie comme un triplet $\mathscr{P} = (\mathscr{X}, \mathscr{T}, (P_\theta)_{\theta\in\Theta})$ où $(\mathscr{X}, \mathscr{T})$ est un espace mesurable, $(P_\theta)_{\theta\in\Theta}$ est une famille de lois de probabilité sur l'espace $(\mathscr{X}, \mathscr{T})$ indexée par un ensemble Θ. L'espace mesurable $(\mathscr{X}, \mathscr{T})$ est appelé *espace des échantillons de l'expérience* $\mathscr{P}$ et l'ensemble Θ est appelé *espace des paramètres de* $\mathscr{P}$.

La notion de modèle statistique a été introduite par Wald (1939, 1950) et reprise dans Blackwell (1951, 1953). Il s'agit d'une abstraction mathématique visant à représenter une expérience concrète. Un exemple simple pourrait être celui d'un sondage. On choisit un échantillon de taille n de la population et on demande à chaque individu s'il a l'intention de voter pour le candidat d'intérêt ou pas. Formellement, ceci amène à poser un modèle de la forme $\mathscr{P} = (\mathscr{X}, \mathscr{T}, (P_\theta)_{\theta\in\Theta})$ où $\mathscr{X} = \{0,1\}^n$ représente les expressions des votes de n individus tirés indépendamment, $\mathscr{T}$ est la tribu des parties de $\mathscr{X}$, $\Theta = [0,1]$ et P_θ est la loi commune aux n variables de Bernoulli indépendantes.

Une expérience étant donnée, on l'utilise pour prendre une décision dont on mesure le risque. Cependant, parfois, on peut modéliser une même réalité à travers plusieurs modèles statistiques différents. Une première proposition sur comment comparer deux modèles statistiques ayant le même espace des paramètres est apparu en 1949 quand Bohnenblust, Shapley, Sherman (1949) ont introduit la définition suivante : "$\mathscr{P}_1$ est *plus informatif que* $\mathscr{P}_2$" si pour toute fonction de perte bornée L et pour toute règle de décision

ρ_2 disponible dans l'expérience $\mathscr{P}_2$ il existe une règle de décision ρ_1 dans l'expérience $\mathscr{P}_1$ telle que

$$R(\mathscr{P}_1, \rho_1, L, \theta) \leq R(\mathscr{P}_2, \rho_2, L, \theta), \quad \forall \theta \in \Theta.$$

Ici on dénote par $R(\mathscr{P}_1, \rho_1, L, \theta)$ et $R(\mathscr{P}_2, \rho_2, L, \theta)$ les *risques statistiques* relatifs aux modèles $\mathscr{P}_1$ et $\mathscr{P}_2$ respectivement (pour une définition précise voir le Chapitre 2, Sous-section 2.3).

Le problème avec une telle définition est qu'elle ne permet pas de comparer n'importe quel couple de modèles statistiques. Une façon d'y remédier est de changer la question qu'on se pose : " Combien d'information statistique sur θ perd-on en utilisant $\mathscr{P}_2$ au lieu de $\mathscr{P}_1$?". Ce changement de perspective a été proposé par le mathématicien français Lucien Le Cam en 1964 et il a amené à introduire la notion de *déficience*, $\delta(\mathscr{P}_1, \mathscr{P}_2)$, de $\mathscr{P}_1$ par rapport à $\mathscr{P}_2$. Cette quantité doit être interprétée comme un indicateur numérique du coût nécessaire pour reconstruire l'expérience $\mathscr{P}_2$ à partir de l'expérience $\mathscr{P}_1$. Avant de se concentrer sur la définition mathématique de la déficience qui nous demandera un long détour, soulignons trois aspects particulièrement intéressants :

- Cette quantité est définie pour n'importe quel couple de modèles $(\mathscr{P}_1, \mathscr{P}_2)$ partageant le même espace des paramètres Θ.

- Pour toute fonction de perte L telle que $0 \leq L \leq 1$, pour toute règle de décision ρ_2 disponible sur Θ en utilisant l'expérience $\mathscr{P}_2$, il existe une règle de décision ρ_1 disponible sur $\mathscr{P}_1$ telle que

  $$R(\mathscr{P}_1, \rho_1, L, \theta) \leq R(\mathscr{P}_2, \rho_2, L, \theta) + \delta(\mathscr{P}_1, \mathscr{P}_2), \quad \forall \theta \in \Theta.$$

  En particulier, si $\delta(\mathscr{P}_1, \mathscr{P}_2) = 0$ et $\delta(\mathscr{P}_2, \mathscr{P}_1) = 0$ alors à toute règle de décision π_2 qui est "bonne" pour $\mathscr{P}_2$ correspond une règle de décision π_1 qui est aussi bonne pour $\mathscr{P}_1$ et réciproquement.

- À la base de la théorie de Le Cam sur la comparaison des expériences statistiques il y a l'idée de passer d'un problème qu'on ne sait pas bien traiter à un autre pour lequel la théorie est bien développée. Souvent en fait, on observe une variable aléatoire X à valeurs dans $\mathscr{X}_1$ ayant pour loi $P_{1,\theta}$, la valeur du paramètre θ est inconnue, et on voudrait estimer θ. Le problème est alors, en utilisant la terminologie de Le Cam lui même, qu' *en général, la famille* $\mathscr{P}_1 = (\mathscr{X}_1, \mathscr{T}_1, (P_{1,\theta})_{\theta \in \Theta})$ *est compliquée. On voudrait alors l'approcher par une famille plus simple* $\mathscr{P}_2 = (\mathscr{X}_2, \mathscr{T}_2, (Q_\theta, \theta \in \Theta))$. Grâce au point précédent, on voit bien que si $\delta(\mathscr{P}_1, \mathscr{P}_2)$ et $\delta(\mathscr{P}_2, \mathscr{P}_1)$ sont simultanément petits, les problèmes de l'estimation de θ dans l'expérience $\mathscr{P}_1$ sont alors réduits (avec une perte d'informations proportionnelle à $\delta(\mathscr{P}_1, \mathscr{P}_2)$) à ceux de l'estimation de θ dans l'expérience plus facile $\mathscr{P}_2$ (*de même que l'on peut approcher localement une courbe par des segments de droites, ou une figure géométrique compliquée à l'aide de triangles*[1]).

D'une façon intuitive, deux expériences $\mathscr{P}_1$ et $\mathscr{P}_2$ sont dites *équivalentes* s'il est possible de reproduire l'une à partir de l'autre et réciproquement via une *transition*, c'est-à-dire, si on dispose d'un mécanisme pour convertir les observations issues de la distribution $P_{1,\theta}$ en observations issues de $P_{2,\theta}$ et réciproquement. Un point clé est dans le sens à donner au mot transition (voir la Section 2.2 du Chapitre 2 pour une définition précise). Des exemples de transition sont donnés par les noyaux markoviens et, en fait, on peut démontrer que sous certaines hypothèses sur les espaces des échantillons, l'ensemble de toutes les transitions possibles coïncide avec l'ensemble des noyaux markoviens $K : (\mathscr{X}_1, \mathscr{T}_1) \to (\mathscr{X}_2, \mathscr{T}_2)$. En particulier, ceci est le cas lorsqu'on considère des modèles statistiques dominés ayant des espaces polonais comme espace des échantillons. Rappelons qu'un noyau markovien $K : (\mathscr{X}_1, \mathscr{T}_1) \to (\mathscr{X}_2, \mathscr{T}_2)$ est une application $K : \mathscr{X}_1 \times \mathscr{T}_2 \to [0,1]$ telle que :

- l'application $x \mapsto K(x, A)$ est $\mathscr{T}_1$-mesurable pour tout $A \in \mathscr{T}_2$;
- l'application $A \mapsto K(x, A)$ est une mesure de probabilité sur $(\mathscr{X}_2, \mathscr{T}_2)$ pour tout $x \in \mathscr{X}_1$.

Si on dispose d'un noyau markovien $K : (\mathscr{X}_1, \mathscr{T}_1) \to (\mathscr{X}_2, \mathscr{T}_2)$ alors à partir de chaque $P_{1,\theta}$ on peut construire une probabilité $\tilde{P}_{2,\theta} = KP_{1,\theta}$ sur $\mathscr{T}_2$, c'est-à-dire

$$\tilde{P}_{2,\theta}(A) = \int K(x, A) P_{1,\theta}(dx), \quad \forall A \in \mathscr{T}_2.$$

On verra dans le Chapitre 2 que, d'une façon générale, une transition est définie comme une application linéaire positive entre des treillis de Banach particuliers (L-spaces), ce qui généralise la définition de noyau de Markov.

Définition 1.1.1. Soit $\mathscr{P}_1$ un modèle statistique dominé et $\mathscr{P}_2$ un modèle statistique avec un espace des échantillons polonais.

La déficience de $\mathscr{P}_1$ par rapport à $\mathscr{P}_2$ est définie par

$$\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_{K \in \mathscr{K}} \sup_{\theta \in \Theta} \|KP_{1,\theta} - P_{2,\theta}\|_{TV}, \tag{1.1}$$

[1] Gaëlle Octavia, SMF-Gazette-118, Octobre 2008

où $\mathscr{K}$ représente l'ensemble de tous les noyaux markoviens de $(\mathscr{X}_1, \mathscr{T}_1)$ vers $(\mathscr{X}_2, \mathscr{T}_2)$ et $\|\cdot\|_{TV}$ désigne la distance en variation totale, i.e.

$$\|KP_{1,\theta} - P_{2,\theta}\|_{TV} := \sup_{A\in\mathscr{T}_2} \big|KP_{1,\theta}(A) - P_{2,\theta}(A)\big|.$$

Définition 1.1.2. La *distance de Le Cam* ou Δ-*écart* entre $\mathscr{P}_1$ et $\mathscr{P}_2$ est ainsi définie :

$$\Delta(\mathscr{P}_1, \mathscr{P}_2) = \max\Big(\delta(\mathscr{P}_1, \mathscr{P}_2), \delta(\mathscr{P}_2, \mathscr{P}_1)\Big).$$

Les modèles statistiques $\mathscr{P}_1$ et $\mathscr{P}_2$ sont dits *équivalents* si $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$.

La définition de distance de Le Cam telle qu'on vient de donner n'éclaircit peut-être complètement pas l'importance statistique d'une telle distance. En effet, un intérêt de la théorie de Le Cam est l'interprétation possible du Δ-écart à l'aide de la théorie de la décision. Pour cela, nous en rappelons ci-dessous le cadre classique.

Un *problème de décision* pour le modèle $\mathscr{P}$ est la donnée d'un espace mesurable $(D, \mathcal{D})$, appelé *espace des décisions*, et d'une fonction $L : D \times \Theta \to [0, \infty)$, appelée *fonction de perte*, telle que, pour tout $\theta \in \Theta$, $L(\cdot, \theta)$ soit mesurable.

Le statisticien observe une valeur $x \in \mathscr{X}$ obtenue à partir de la mesure de probabilité P_θ. Il ne connaît pas la valeur de θ et sa tâche est de sélectionner une décision $z \in D$. Pour l'accomplir, il choisit une mesure de probabilité $\pi(x, \cdot)$ sur D et il tire un point de D au hasard selon la loi $\pi(x, \cdot)$. S'il a choisi z quand la vraie distribution de x est P_θ, il subit une perte $L(\theta, z)$.

La fonction $\pi : x \to \pi(x, \cdot)$ est appelée *règle de décision randomisée* dans le modèle $\mathscr{P}$. Autrement dit, π est un noyau markovien de $(\mathscr{X}, \mathscr{T})$ vers $(D, \mathcal{D})$. Notons $\mathcal{L}(D, \mathcal{D})$ l'ensemble de toutes les fonctions de perte L qui satisfont $0 \leq L(z, \theta) \leq 1, \forall z \in D, \forall \theta \in \Theta$ et avec $\Pi(\mathscr{P})$ l'ensemble des règles de décision randomisées dans le modèle $\mathscr{P}$.

On appelle *risque* de la règle de décision randomisée $\pi \in \Pi(\mathscr{P})$ en θ associé à la fonction de perte $L \in \mathcal{L}(D, \mathcal{D})$ la quantité donnée par

$$R(\mathscr{P}, \pi, L, \theta) = \int_{\mathscr{X}} \bigg(\int_D L(z, \theta)\pi(x, dz)\bigg) P_\theta(dx).$$

Proposition 1.1.3 (Le Cam (1986))**.** *Soit $\varepsilon > 0$ fixé. Deux modèles statistiques $\mathscr{P}_1$ et $\mathscr{P}_2$ vérifient la relation $\delta(\mathscr{P}_1, \mathscr{P}_2) \leq \varepsilon$ si et seulement si : pour tout problème de décision avec fonction de perte $L \in \mathcal{L}(D, \mathcal{D})$ et pour toute fonction de décision $\pi \in \Pi(\mathscr{P}_2)$, il existe une fonction de décision $\pi^* \in \Pi(\mathscr{P}_1)$ telle que*

$$R(\mathscr{P}_1, \pi^*, L, \theta) \leq R(\mathscr{P}_2, \pi, L, \theta) + \varepsilon, \quad \theta \in \Theta.$$

Soulignons que cette caractérisation concerne la déficience $\delta(\mathscr{P}_1, \mathscr{P}_2)$ de $\mathscr{P}_1$ par rapport à $\mathscr{P}_2$, et que δ n'est pas symétrique. Par conséquent, si pour l'expression symétrisée (le Δ-écart) on a $\Delta(\mathscr{P}_1, \mathscr{P}_2) \leq \varepsilon$, alors les risques des procédures d'estimation dans $\mathscr{P}_2$ sont les mêmes que ceux dans $\mathscr{P}_1$, à ε près, et vice-versa.

On peut maintenant donner une définition de la déficience en termes de règles de décision. Plus précisément on a la proposition suivante :

Proposition 1.1.4 (Le Cam (1986))**.** *La définition de déficience donnée par* (1.1) *est équivalente à :*

$$\delta(\mathscr{P}_1, \mathscr{P}_2) = \sup \sup_{L \in \mathcal{L}(D,\mathcal{D})} \inf_{\pi_1 \in \Pi(\mathscr{P}_1)} \sup_{\pi_2 \in \Pi(\mathscr{P}_2)} \sup_{\theta \in \Theta} |R(\mathscr{P}_1, \pi_1, L, \theta) - R(\mathscr{P}_2, \pi_2, L, \theta)|,$$

où le premier sup est pris sur tous les espaces de décision $(D, \mathcal{D})$.

1.1.2 État de l'art

Les premiers résultats d'équivalence asymptotique dans un contexte non paramétrique, c'est-à-dire où Θ est un ensemble infini-dimensionnel, datent de 1996 et sont dus à Brown, Low (1996) et Nussbaum (1996). Dans Brown, Low (1996) les auteurs ont établi un résultat d'équivalence asymptotique entre un modèle de régression non paramétrique

$$Y_i = f(x_i) + \sigma(x_i)\xi_i, \quad i = 1, \dots, n, \ (\xi_i) \text{ i.i.d. gaussiennes centrées et réduites,}$$

et un modèle de bruit blanc gaussien

$$dY_t = f(t)dt + \frac{\sigma(t)}{\sqrt{n}} dW_t, \quad t \in [0,1], \quad (W_t) \text{ mouvement Brownien standard.}$$

L'équivalence est établie en supposant que le coefficient de diffusion $\sigma(\cdot)$ est connu et que f appartient à $\mathscr{F}$, une classe fonctionnelle constituée de fonctions suffisamment lisses (i.e. $\Theta = \mathscr{F}$). Au cours des années, plusieurs extensions de ce résultat ont été proposées. Dans Brown et al. (2002a), les auteurs traitent le cas où les covariables x_i sont aléatoires, tandis que dans Rohde (2004) des améliorations dans l'approximation sont discutées. Le cas d'une variance σ inconnue est traité dans Carter (2007). Pour des résultats multidimensionnels, nous pouvons citer Carter (2006a) ; Reiß (2008). L'article de Grama, Nussbaum (1998) est le premier à considérer le cas de bruits non forcément gaussiens, mais appartenant à une famille de lois exponentielles ; les auteurs ont ultérieurement développé ce résultat dans Grama, Nussbaum (2002) pour prendre en compte des bruits avec des densités quelconques à condition qu'elles soient suffisamment lisses. Dans Carter (2009), l'auteur considère les effets d'une corrélation entre les bruits : il s'avère que l'expérience limite n'est plus un modèle de bruit blanc gaussien, mais elle est plus proche d'un mouvement brownien fractionnaire. Si, par contre, on s'intéresse à des bruits dont la densité peut être discontinue, alors l'expérience limite est encore différente comme prouvé dans l'article de Meister, Reiß (2013). En ce qui concerne les résultats d'équivalence pour les modèles à densité le travail fondateur est celui de Nussbaum (1996). Dans cet article, Nussbaum a démontré l'équivalence asymptotique entre l'expérience

$$(X_i)_{1\leq i\leq n} \quad \text{variables aléatoires i.i.d. de densité} f \text{ définie sur } [0,1]$$

et le modèle de bruit blanc gaussien

$$dY_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{n}}dW_t, \quad t\in[0,1], \quad (W_t) \text{ mouvement Brownien standard,}$$

en ce qui concerne l'estimation de f, qu'on suppose appartenir à une certaine classe fonctionnelle $\mathscr{F}$. La preuve de ce résultat telle qu'elle a été présentée dans Nussbaum (1996) ne permet pas d'expliciter les transitions en jeu, mais nous pouvons citer Brown et al. (2004a) ; Carter (2002) pour des extensions du travail de Nussbaum où, en particulier, des noyaux de Markov sont construits explicitement. Dans Jähnisch, Nussbaum (2003) le cas de variables aléatoires $(X_i)_i$ indépendantes mais pas de même loi a été traité. Un autre domaine riche de résultats d'équivalence au sens de Le Cam est celui des modèles de diffusion. Dans ce contexte, le premier résultat est dû à Milstein, Nussbaum (1998). Ils ont prouvé l'équivalence asymptotique entre un modèle de diffusion unidimensionnel en petite variance (i.e. avec un coefficient de diffusion petit) :

$$dy_t = f(y_t)dt + \varepsilon dW_t, \quad y_0 = 0 \quad t\in[0,1], \quad (W_t) \text{ mouvement Brownien standard}$$

et le schéma d'Euler associé, avec un pas d'échantillonnage $1/n$:

$$Z_0 = 0, \quad Z_i = Z_{i-1} + \frac{f(Z_{i-1})}{n} + \frac{\varepsilon}{\sqrt{n}}\xi_i, \quad i = 1,\dots,n,$$

avec les ξ_i i.i.d. de loi gaussienne centrée et réduite et $\varepsilon \to 0$. Ils ont aussi prouvé l'équivalence entre l'observation continue de $(y_t)_{t\in[0,1]}$ et celle discrète de$\big(y_{\frac{i}{n}}\big)_{0\leq i\leq n}$. D'autres résultats d'équivalence entre des processus de diffusion et les schémas d'Euler correspondants sont démontrés dans Dalalyan, Reiß (2006) (voir Sous-section 4.5) et Genon-Catalot, Laredo (2014). En restant dans le thème des modèles de diffusion nous pouvons citer l'article de Delattre, Hoffmann (2002), dans lequel les auteurs ont étudié le problème de l'équivalence pour des diffusions de la forme

$$X_t = x_0 + \int_0^t f(X_s)ds + W_t, \quad t\geq 0,$$

sous des hypothèses pour lesquelles $(X_t)_{t\geq 0}$ est un processus de Markov récurrent nul au sens de Harris (voir, par exemple, Revuz, Yor (1999)). Dans le même esprit, on peut aussi citer les résultats de Dalalyan, Reiß (2006, 2007a). Finalement, un résultat d'équivalence pour des diffusions en petite variance et observées dans une fenêtre temporelle aléatoire est celui de Genon-Catalot, Laredo, Nussbaum (2002). Un autre résultat remarquable concernant l'équivalence pour des variables aléatoires dépendantes est dû à Grama, Neumann (2006), qui ont prouvé l'équivalence asymptotique entre un modèle d'autorégression non paramétrique et une régression non paramétrique. Par ailleurs, des résultats d'équivalence existent aussi pour des problèmes inverses dans la régression et le modèle de bruit blanc (Meister (2011)), pour l'estimation spectrale d'une densité, (Golubev, Nussbaum, Zhou (2010)) et pour l'estimation du coefficient de diffusion à partir d'observations en haute fréquence d'une martingale continue plus un bruit, (Reiß (2011)).

1.2 Résultats principaux de la thèse

Nous présentons maintenant les résultats principaux qui sont contenus dans cette thèse. Il s'agit principalement de résultats d'équivalence asymptotique au sens de Le Cam (voir les Sous-sections 1.2.1–1.2.3) et d'une majoration en distance L_1 entre les lois de processus additifs (voir la Sous-section 1.2.4) qui a fait l'objet d'une publication (Étoré, Mariucci (2014)).

Dans la Sous-section 1.2.1 nous présentons deux résultats pour des processus à sauts. Le premier résultat présenté traite de l'équivalence, en ce qui concerne l'estimation de la dérive, entre un processus additif observé à haute fréquence et un modèle de bruit blanc gaussien ; ce résultat a donné lieu à la publication Mariucci (2015b). Dans la deuxième partie de la Sous-section 1.2.1 nous introduisons un résultat d'équivalence concernant la densité de Lévy : nous avons démontré qu'estimer la densité de Lévy issue d'un processus de Lévy à sauts purs (observé d'une manière continue ou discrète) est équivalent à estimer la dérive d'un certain modèle de bruit blanc gaussien ; ce résultat est présenté dans l'article Mariucci (2015d). Dans la Sous-section 1.2.2 nous discutons brièvement une extension du résultat de Nussbaum (1996), sur l'équivalence pour des modèles à densité. Ce résultat est détaillé dans le Chapitre 5 qui est basé sur l'article Mariucci (2015a). Finalement, dans la Sous-section 1.2.3 nous discutons un résultat d'équivalence qui traite le lien entre des processus de diffusion en petite variance et le schéma d'Euler correspondant lorsque le coefficient de diffusion est non constant ; ce résultat a fait l'objet de la publication Mariucci (2015c).

1.2.1 Équivalence asymptotique pour des modèles à sauts

Les processus de Lévy sont des processus à accroissements stationnaires et indépendants, à trajectoires presque sûrement continues à droite ayant une limite à gauche (càdlàg). La classe de ces processus est extrêmement riche, les représentants les plus connus étant le processus de Poisson ou de Poisson composé, le mouvement Brownien, le processus de Cauchy et, plus généralement, les processus stables. Ils sont utilisés pour modéliser des situations où des changements soudains peuvent se produire. Pour cela, ils sont largement utilisés en finance, mais aussi dans d'autres domaines comme, par exemple, les télécommunications, la statistique des valeurs extrêmes, la mécanique quantique et la biologie. Ils sont les éléments de base permettant de construire des modèles stochastiques en temps continu avec des sauts. À titre d'exemple, on pourrait citer les "exponential Lévy models", "hyperbolic Lévy motions", "time changed Lévy processes" ou "Stochastic volatility models". Pour cette raison, l'inférence statistique pour des processus de Lévy a suscité une attention considérable ces dernières années, devenant un sujet de grand intérêt aussi bien pour les théoriciens que pour les praticiens.

D'un point de vue mathématique, la loi d'un processus de Lévy $X = (X_t)_{t\geq 0}$ est uniquement déterminée par trois paramètres : la *dérive* $\gamma \in \mathbb{R}$, la *variance* (ou *coefficient de diffusion*) $\sigma^2 \geq 0$ et la *mesure de Lévy* ν.

Cependant, l'hypothèse de stationnarité les rend souvent peu flexibles, par conséquent on préfère parfois modéliser plutôt avec des processus additifs, c'est-à-dire des processus analogues à ceux de Lévy mais ayant une fonction caractéristique de la forme

$$\mathbb{E}\big[e^{iuX_t}\big] = \exp\Big(iu\int_0^t f(r)dr - \frac{u^2}{2}\int_0^t \sigma^2(r)dr - \int_{\mathbb{R}}(1 - e^{iuy} + iuy\mathbb{I}_{|y|\leq 1})\nu_t(dy)\Big), \quad (1.2)$$

où $f(\cdot)$ et $\sigma^2(\cdot)$ appartiennent à $L_1(\mathbb{R})$ et ν_t est une mesure positive sur $\mathbb{R}$ qui satisfait

$$\nu_t(\{0\}) = 0 \text{ et } \int_{\mathbb{R}}(y^2 \wedge 1)\nu_t(dy) < \infty, \quad \forall t \geq 0,$$

ainsi qu'une condition de monotonicité : si $0 < s \leq t$ alors $\nu_s(A) \leq \nu_t(A)$, pour tout $A \in \mathscr{B}(\mathbb{R})$ (voir Sato (1999), Théorème 9.8).

La représentation donnée en (1.2) en termes du triplet $(f(t), \sigma^2(t), \nu_t)_{t\geq 0}$ est unique et cette dernière est appelée *caractéristique locale* du processus X alors que ν_t comme ci-dessous est dite *mesure de Lévy*, pour tout t. Dans le cas où $f(\cdot) \equiv \gamma$ et $\sigma(\cdot) \equiv \sigma$ sont des fonctions constantes et $\nu_t = \nu$ pour tout t, le processus X en (1.2) est un processus de Lévy de *triplet* (γ, σ^2, ν).

La théorie de l'estimation des composantes du triplet caractéristique pour des processus de Lévy observés d'une façon continue est assez bien développée et plusieurs estimateurs ont été proposés. Cependant, pour de nombreux problèmes pratiques il n'est pas possible, ni économique, d'observer toute une trajectoire d'un processus et on est obligé de travailler avec des observations discrètes, c'est-à-dire avec un vecteur d'observations

$$(X_{t_0}, X_{t_1}, \dots, X_{t_n}),$$

où $0 = t_0 < t_1 < \dots < t_n = T_n$ et $\Delta_n := \max_{1 \leq i \leq n}\{|t_i - t_{i-1}|\}$ est petit. Il est désormais classique dans l'inférence statistique pour les processus en temps continu et observés d'une façon discrète, de distinguer deux points de vue. Un premier, est celui de la basse fréquence : le pas de discrétisation Δ_n est maintenu fixe tandis que le nombre n d'observations tend vers l'infini. On parle de haute fréquence quand le pas de discrétisation Δ_n tend vers 0 lorsque n tend vers l'infini. Cette dernière approche a joué un rôle central dans la littérature récente sur l'estimation non paramétrique pour les processus de Lévy ou additifs et elle sera le point de vue adopté dans cette thèse.

Or, il semblerait naturel de penser que si Δ_n tend de façon suffisamment rapide vers 0 lorsque le nombre d'observations n tend vers l'infini, alors il n'y a pas "trop de perte d'information" quand on passe d'un modèle associé à l'observation continue de X à un modèle associé aux observations discrètes de X. Ceci peut être formalisé en utilisant la théorie de Le Cam et c'est ce que nous faisons dans les Chapitres 3 et 4.

Plus précisément, dans un premier temps nous nous sommes intéressés au problème de l'estimation non paramétrique de la dérive d'un processus additif ayant pour caractéristique locale $(f(t), \sigma^2(t), \nu_t)_{t\geq 0}$. Ensuite, nous nous sommes focalisés sur le problème de l'estimation de la densité de Lévy relative à un processus de Lévy à sauts purs. Dans les deux cas, on a supposé ne disposer que d'observations à haute fréquence du processus sous-jacent X. Si on dénote par $\theta \in \Theta$ un paramètre (i.e. Θ est l'espace lié à la dérive dans le premier cas et à la densité de Lévy dans le deuxième) nous nous sommes posés essentiellement deux questions :

(1) Combien d'informations sur le paramètre θ perdons-nous en observant $(X_{t_i})_{i=0}^n$ à la place de $(X_t)_{t\in[0,T_n]}$?

(2) Peut-on construire un modèle plus simple d'un point de vue mathématique, mais équivalent du point de vue de l'information sur θ, à partir de l'observation de $(X_{t_i})_{i=0}^n$?

Dans les deux paragraphes qui suivent nous allons donner une réponse aux questions (1) et (2) en utilisant la théorie de Le Cam sur les expériences statistiques.

Estimation de la dérive

En ce qui concerne notre contribution aux problèmes de l'équivalence asymptotique pour des modèles à sauts, nous avons commencé par considérer l'expérience statistique associée à l'observation discrète à haute fréquence d'un processus gaussien et d'un processus de Poisson composé inhomogène en temps. Plus précisément, nous nous sommes focalisés sur l'étude d'un processus de la forme

$$X_t = \eta + \int_0^t f(s)ds + \int_0^t \sigma_n(s)dW_s + \sum_{i=1}^{N_t} Y_i, \quad t \in [0, T_n], \tag{1.3}$$

où :

- η est une condition initiale ;
- $W = (W_t)_{t\geq 0}$ est un mouvement Brownien standard ;
- $N = (N_t)_{t\geq 0}$ est un processus de Poisson, inhomogène en temps, avec une intensité notée par $\lambda(\cdot)$, indépendant de W ;
- $(Y_i)_{i\geq 1}$ est une suite de variables aléatoires réelles indépendantes et identiquement distribuées de loi G, indépendante de W et N.

Ici, le paramètre inconnu à estimer est la dérive $f(\cdot)$ qui est supposée appartenir à une certaine classe non paramétrique $\mathscr{F}$. Le coefficient de diffusion $\sigma_n(\cdot)$ est supposé connu contrairement aux paramètres liés à la partie discontinue de $X = (X_t)_{t\geq 0}$. Plus précisément, l'intensité $\lambda(\cdot)$ et la distribution G de Y_1, ne sont pas supposées connues et elles appartiennent à certaines classes non paramétriques Λ et $\mathscr{G}$, respectivement. Nous disposons de $n+1$ observations $X_{t_0}, X_{t_1}, \dots, X_{t_n}$ avec $\Delta_n \to 0$ lorsque $n \to \infty$. Concernant l'horizon temporel T_n, il peut être soit fini, $T_n \equiv T$, soit convergent vers l'infini lorsque $n \to \infty$. Nous nous intéressons au problème de l'estimation de la dérive à partir des observations discrètes (X_{t_i}). Pour garantir l'identifiabilité de $f(\cdot)$, si $T_n \equiv T$, nous nous plaçons en petite variance, c'est-à-dire, $\sigma_n(\cdot) = \varepsilon_n\sigma(\cdot)$ avec $\varepsilon_n \to 0$ quand $n \to \infty$. Si $T_n \to \infty$, la classe $\mathscr{F}$ sera une sous-classe de fonctions périodiques.

Tout d'abord, nous allons reformuler le problème en termes de modèles statistiques. La notation introduite ci-dessous restera valable tout au long de ce chapitre introductif.

Notation 1.2.1.

- *Nous notons* $C = C([0,\infty],\mathbb{R})$ *l'espace des fonctions réelles continues sur* $[0,\infty]$. *Soit* $x : C \to C$ *le* processus canonique, *c'est-à-dire, le processus tel que*
$$\forall \omega \in C, \quad x_t(\omega) = \omega_t \quad \forall t \geq 0.$$
Nous notons $\mathscr{C}^0$ *la plus petite tribu qui rend* x_s *mesurable,* $s \geq 0$*. Ensuite, pour tout* $t \geq 0$*, soit* $\mathscr{C}_t^0$ *la plus petite tribu qui rend* x_s *mesurable, pour tout* $s \in [0,t]$*. Enfin, nous définissons* $\mathscr{C}_t := \cap_{s>t}\mathscr{C}_s^0$ *et* $\mathscr{C} = \sigma(\mathscr{C}_t; t > 0)$*.*

- *Nous notons* $D = D([0,\infty],\mathbb{R})$ *l'espace des fonctions* $\omega : [0,\infty] \to \mathbb{R}$ *qui sont continues à droite et qui admettent une limite à gauche en chaque point (càdlàg). Avec un léger abus de notation, nous notons également* x *le processus canonique sur* D *:*
$$\forall \omega \in D, \quad x_t(\omega) = \omega_t \quad \forall t \geq 0.$$
Nous notons $\mathscr{D}_t$ *et* $\mathscr{D}$ *les tribus engendrées par* $\{x_s : 0 \leq s \leq t\}$ *et* $\{x_s : 0 \leq s < \infty\}$*, respectivement. L'espace mesurable* $(D,\mathscr{D})$ *est l'*espace de Skorokhod.

- *Soit* X *un processus additif défini sur l'espace probabilisé* $(\Omega,\mathscr{A},\mathbb{P})$ *avec une caractéristique locale* $(f(t),\sigma^2(t),\nu_t)_{t\geq 0}$*. Ce processus induit une mesure de probabilité sur* $(D,\mathscr{D})$ *que l'on notera* $P^{(f,\sigma^2,\nu)}$*. Notons que le processus canonique* x *défini sur* $(D,\mathscr{D},P^{(f,\sigma^2,\nu)})$ *est un processus additif de même loi que* $(X,\mathbb{P})$ *(i.e. la caractéristique locale de* x *sous* $P^{(f,\sigma^2,\nu)}$ *est* $(f(t),\sigma^2(t),\nu_t)_{t\geq 0}$*). De plus, nous noterons* $P_t^{(f,\sigma^2,\nu)}$ *la restriction de* $P^{(f,\sigma^2,\nu)}$ *à la tribu* $\mathscr{D}_t$*.*

- *Pour toute fonction* ω *dans* D*, nous noterons* $\Delta\omega_r$ *le saut au temps* r *et* ω^c *ou* ω^d *sa partie continue ou discontinue, respectivement :*
$$\Delta\omega_r = \omega_r - \lim_{s\uparrow r}\omega_s, \ \ \omega_t^d = \sum_{r\leq t}\Delta\omega_r, \ \ \omega_t^c = \omega_t - \omega_t^d.$$

- *Soit* X *un processus additif avec une caractéristique locale* $(f(t),\sigma^2(t),\nu_t)_{t\geq 0}$*. Pour tout* $0 = t_0 < t_1 < \cdots < t_n$*, nous noterons* $Q_n^{(f,\sigma_n^2,\nu)}$ *la loi du vecteur* $(X_{t_0},X_{t_1},\dots,X_{t_n})$ *sur* $(\mathbb{R}^{n+1},\mathscr{B}(\mathbb{R}^{n+1}))$*.*

Ici, l'espace des paramètres est $\mathscr{F}$ et les modèles statistiques que nous considérons sont les suivants :
$$\mathscr{P}_n = \big(D,\mathscr{D},\{P_{T_n}^{(f,\sigma_n^2,\lambda G)} : f \in \mathscr{F}\}\big), \tag{1.4}$$
$$\mathscr{Q}_n = \big(\mathbb{R}^{n+1},\mathscr{B}(\mathbb{R}^{n+1}),\{Q_n^{(f,\sigma_n^2,\lambda G)} : f \in \mathscr{F}\}\big), \tag{1.5}$$

où par λG nous notons la mesure de Lévy $\nu_t(A) = \lambda(t)G(A)$, pour tout t, pour tout A.

Finalement, nous introduisons le modèle gaussien qui apparaîtra dans l'énoncé de nos principaux résultats. Pour cela, de manière cohérente avec la notation précédente, nous écrirons $P_{T_n}^{(f,\sigma_n^2,0)}$ pour désigner la loi induite sur $(C,\mathscr{C}_{T_n})$ par le processus stochastique :

$$dy_t = f(t)dt + \sigma_n(t)dW_t, \quad y_0 = 0, \quad t \in [0,T_n]. \tag{1.6}$$

Nous posons :

$$\mathscr{W}_n = \big(C, \mathscr{C}_{T_n}, (P_{T_n}^{(f,\sigma_n^2,0)} : f \in \mathscr{F})\big). \tag{1.7}$$

Nous avons déjà mentionné que les équivalences asymptotiques peuvent être utilisées pour réduire les problèmes d'estimation d'un modèle à ceux d'un autre modèle plus simple. C'est le cas ici : le modèle associé à l'observation discrète ou continue de X comme dans (1.3) s'est révélé être équivalent à celui de (1.6), qui a fait l'objet de nombreuses études (voir Ibragimov, Has'minskii 1981). Par exemple, considérons les deux situations suivantes :

- T_n est fixé et $\sigma_n(\cdot) = \varepsilon_n\sigma(\cdot)$ avec $\varepsilon_n \to 0$;
- T_n tend vers l'infini et $\sigma_n(\cdot)$ est fixé ; dans ce cas, nous demandons également que les éléments de $\mathscr{F}$ satisfassent des hypothèses de périodicité.

Dans les deux cas, une estimation consistante de $f \in \mathscr{F}$ est possible. Nos résultats d'équivalence ne s'appuient pas sur des hypothèses comme celles-ci, mais ils s'appliquent aussi à ces cas. En effet, l'existence d'une équivalence pour une classe $\mathscr{F}$ entraîne la même équivalence pour toute sous-classe de $\mathscr{F}$.

Nous énonçons ci-dessous notre résultat principal dans le cas où $\mathscr{F}$ est une classe de fonctions α-Hölder, uniformément bornées sur $\mathbb{R}$, i.e. il existe $B < \infty$, $M < \infty$ et $\alpha \in (0,1]$ tels que

$$|f(x)| \leq B \quad \text{et} \quad |f(x) - f(y)| \leq M|x-y|^\alpha, \quad \forall x,y \in \mathbb{R}.$$

Pour un énoncé plus général, voir la Sous-section 3.2.4.

Théorème 1.2.2. *Soit $\mathscr{F}$ une sous-classe de fonctions α-Hölder, uniformément bornées sur $\mathbb{R}$. Soit $\sigma_n(\cdot) = \varepsilon_n\sigma(\cdot)$ telle que $0 < m_\sigma \leq \sigma(\cdot) \leq M_\sigma < \infty$ avec une dérivée $\sigma'(\cdot)$ dans $L_\infty(\mathbb{R})$. Supposons que l'une des deux conditions suivantes soit satisfaite :*

- *$T_n \equiv T < \infty$, $\varepsilon_n \to 0$ et il existe une constante $0 < L_2 < \infty$ telle que pour tout $\lambda \in \Lambda$, $\|\lambda\|_{L_2([0,T])} < L_2$.*

- $T_n \to \infty$, $\varepsilon_n \equiv 1$ *et il existe deux constantes* $0 < L_1, L_2 < \infty$ *telles que pour tout* $\lambda \in \Lambda$, $\|\lambda\|_{L_1(\mathbb{R})} < L_1$, $\|\lambda\|_{L_2(\mathbb{R})} < L_2$.

Alors les expériences $\mathscr{P}_n$, $\mathscr{Q}_n$ *et* $\mathscr{W}_n$ *définies par* (1.4), (1.5) *et* (1.7)*, respectivement, sont asymptotiquement équivalentes dès que l'une des deux conditions suivantes est satisfaite :*

1. $\mathscr{G}$ *est une sous-classe de distributions discrètes avec support sur* $\mathbb{Z}$ *: dans ce cas, une majoration pour la vitesse de convergence est donnée par* $O\Big(\sqrt{\Delta_n} + T_n\Delta_n^{2\alpha}\varepsilon_n^{-2} + T_n\Delta_n\Big)$.
2. $\mathscr{G}$ *est une sous-classe de distributions absolument continues par rapport à la mesure de Lebesgue sur* $\mathbb{R}$ *avec des densités uniformément bornées sur un voisinage fixé de* 0 *: dans ce cas, une majoration pour la vitesse de convergence est donnée par* $O\Big(\sqrt[4]{\Delta_n} + T_n\Delta_n^{2\alpha}\varepsilon_n^{-2} + T_n\Delta_n\Big)$.

La preuve du Théorème 1.2.2, qui est à notre connaissance le premier résultat d'équivalence asymptotique pour des processus à sauts, est constructive. Les idées à la base de la preuve sont les suivantes. Observer les X_{t_i} est clairement équivalent à observer les accroissements $X_{t_i} - X_{t_{i-1}}$, $i = 1, \dots, n$. Or, vu que X est un processus additif, les accroissements sont indépendants. Ceci nous permet de raisonner accroissement par accroissement. De plus, grâce à la forme particulière de X, nous savons que la loi de $X_{t_i} - X_{t_{i-1}}$ peut s'écrire comme la convolution entre une variable aléatoire gaussienne $\mathscr{N}(m_i, \sigma_i^2)$ et la somme aléatoire $\sum_{j=1}^{P_i} Y_j$: avec $m_i = \int_{t_{i-1}}^{t_i} f(s)ds$, $\sigma_i^2 = \int_{t_{i-1}}^{t_i} \sigma_n^2(s)ds$ et P_i une variable aléatoire de Poisson de moyenne $\lambda_i = \int_{t_{i-1}}^{t_i} \lambda(s)ds$. Une première étape est d'utiliser le fait que si $\Delta_n \to 0$ alors on peut se ramener à un modèle où, entre les instants t_{i-1} et t_i, il y a au plus un saut. En termes mathématiques, cela se traduit en passant de la convolution $\mathscr{N}(m_i, \sigma_i^2) * \sum_{j=1}^{P_i} Y_j$ à $\mathscr{N}(m_i, \sigma_i^2) * \varepsilon_i Y_i$, où les ε_i, $i = 1, \dots, n$ sont des variables aléatoires indépendantes de loi Bernoulli de paramètres $\lambda_i e^{-\lambda_i}$, $i = 1, \dots, n$. Comme nous nous intéressons à l'estimation de f et que cette quantité n'intervient que dans la partie gaussienne de la convolution $\mathscr{N}(m_i, \sigma_i^2) * \varepsilon_i Y_i$, il est naturel de penser qu'observer les $\mathscr{N}(m_i, \sigma_i^2) * \varepsilon_i Y_i$ soit asymptotiquement équivalent à observer les $\mathscr{N}(m_i, \sigma_i^2)$ seulement. Pour le démontrer, nous construisons d'une manière explicite les noyaux qui réalisent l'équivalence asymptotique (voir la Sous-section 3.3.2). Quelle que soit la nature de $G \in \mathscr{G}$ (i.e. G à support dans $\mathbb{Z}$ ou absolument continue par rapport à la mesure de Lebesgue), les noyaux construits ne dépendent pas de $\lambda \in \Lambda$. De plus, si la distribution G est concentrée sur $\mathbb{Z}$, le noyau proposé est également indépendant de σ. Ceci, combiné au travail de Carter (2007), nous permet d'étendre notre résultat d'équivalence au cas où la fonction σ est un paramètre de nuisance supposé inconnu. Une fois que l'équivalence asymptotique entre les X_i et les variables gaussiennes $\mathscr{N}(m_i, \sigma_i^2)$ est prouvée, on déduit aisément l'équivalence avec le modèle de bruit blanc (1.7), en utilisant des techniques très proches de celles utilisées dans Brown, Low (1996).

L'équivalence entre $\mathscr{P}_n$ et $\mathscr{W}_n$ est obtenue par l'exhaustivité de l'application S qui à chaque fonction càdlàg associe sa partie continue et du fait que la loi image de S sous $P_{T_n}^{(f,\sigma_n^2,\lambda G)}$ est $P_{T_n}^{(f,\sigma_n^2,0)}$.

Estimation de la densité de Lévy

La dynamique des sauts d'un processus de Lévy est entièrement dictée par sa densité de Lévy. Si on suppose continue cette dernière, sa valeur en un point x_0 représente la fréquence des sauts de taille proche de x_0 qui se produisent par unité de temps. Concrètement, si X est un processus de Lévy à sauts purs avec une densité de Lévy f, alors

$$\int_A f(x)dx = \frac{1}{t}\mathbb{E}\Big[\sum_{s\leq t}\mathbb{I}_A(\Delta X_s)\Big],$$

pour tout borélien A et $t>0$. Ici, $\Delta X_s \equiv X_s - X_{s^-}$ désigne l'amplitude du saut de X au temps s et $\mathbb{I}_A$ la fonction caractéristique. Ainsi, la mesure de Lévy

$$\nu(A) := \int_A f(x)dx,$$

est le nombre moyen des sauts (par unité de temps) dont l'amplitude appartient à l'ensemble A. Comprendre le comportement des sauts nécessite donc d'estimer la mesure de Lévy. Plusieurs travaux récents ont traité ce problème, voir par exemple Belomestny et al. (2015) pour une présentation détaillée.

Dans le Chapitre 4 nous présentons un résultat d'équivalence asymptotique entre les modèles associés à l'observation continue ou discrète d'un processus à sauts purs avec densité de Lévy f et un modèle de bruit blanc gaussien ayant comme dérive $\sqrt{f}$. Les observations discrètes sont échantillonnées avec un pas constant de sorte que $t_i - t_{i-1} = \Delta_n \to 0$, pour tout $i = 1, \dots, n$. De plus, nous supposons que le temps d'observation $T_n \to \infty$ lorsque $n \to \infty$, ce qui permet l'identification de la partie à saut dans la limite. Une telle approche a joué un rôle central dans la littérature récente sur l'estimation non paramétrique pour les processus de Lévy (voir par exemple Bec, Lacour (2012) ; Comte, Genon-Catalot (2010, 2011) ; Duval (2013) ; Figueroa-López (2009)).

Plus en détail, nous considérons le problème de l'estimation de la densité de Lévy $f := \frac{d\nu}{d\nu_0} : I \to \mathbb{R}$, à partir d'un processus de Lévy à sauts purs X de mesure de Lévy ν, observé d'une manière continue ou discrète. Ici ν_0 est une mesure de Lévy fixée et ν peut être une mesure de Lévy infinie si ν_0 l'est ; $I \subseteq \mathbb{R}$ désigne un intervalle éventuellement infini. Dans le cas où ν est à variation finie on peut écrire :

$$X_t = \sum_{0<s\leq t} \Delta X_s \tag{1.8}$$

ou, de manière équivalente, X a une fonction caractéristique de la forme :

$$\mathbb{E}\big[e^{iuX_t}\big] = \exp\bigg(-t\bigg(\int_I (1-e^{iuy})\nu(dy)\bigg)\bigg).$$

Nous supposons que f appartient à un ensemble $\mathscr{F}$, en général non paramétrique. Nous noterons $\mathscr{P}_n^{\nu_0}$ le modèle statistique associé à l'observation continue de X jusqu'à l'instant T_n et $\mathscr{Q}_n^{\nu_0}$ celui associé à l'observation des données discrètes $(X_{t_i})_{i=1}^n$. Bien que tous les résultats du Chapitre 4 soient valables dans le cas général où ν est une mesure de Lévy infinie (à variation finie ou pas), pour des raisons de clarté, nous énonçons ici le théorème principal dans le cas particulier où X est un processus de Poisson composé avec une densité de Lévy différentiable dont la dérivée est uniformément γ-Hölder. Ce cas est spécial pour trois raisons différentes : tout d'abord parce que le théorème s'applique aussi en présence de mesures de Lévy infinies avec une variation éventuellement infinie ; en second lieu parce que la classe fonctionnelle $\mathscr{F}$ peut ne pas être contenue dans une classe de fonctions hölderiennes continûment dérivables et enfin, parce qu'ici I est un intervalle compact par hypothèse, mais une version du théorème reste également vraie pour les intervalles non compacts $I \subseteq \mathbb{R}$.

Théorème 1.2.3. *Soit $I \subseteq \mathbb{R}$ un intervalle compact de $\mathbb{R}$. Pour tous $\gamma \in (0,1]$ et K, κ, M constants strictement positifs, nous considérons la classe fonctionnelle*

$$\mathscr{F}_{(\gamma,K,\kappa,M)} = \Big\{f \in C^1(I) : \kappa \leq f(x) \leq M,\ |f'(x) - f'(y)| \leq K|x-y|^\gamma,\ \forall x,y \in I\Big\}.$$

On suppose que $\mathscr{F} \subseteq \mathscr{F}_{(\gamma,K,\kappa,M)}$. Soit $\mathscr{Q}_n$ (respectivement $\mathscr{P}_n$) le modèle statistique associé à l'observation discrète (respectivement continue) d'un processus de Poisson composé avec densité de Lévy f, par rapport à la mesure de Lebesgue. Finalement, soit $\mathscr{W}_n$ le modèle de bruit blanc gaussien :

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n}}dW_t, \quad t \in I.$$

Alors, les trois modèles $\mathscr{P}_n$, $\mathscr{Q}_n$ et $\mathscr{W}_n$ sont asymptotiquement équivalents :

$$\lim_{n\to\infty} \Delta(\mathscr{Q}_n, \mathscr{W}_n) = 0, \quad \lim_{n\to\infty} \Delta(\mathscr{P}_n, \mathscr{W}_n) = 0 \ \text{ et } \ \lim_{n\to\infty} \Delta(\mathscr{P}_n, \mathscr{Q}_n) = 0. \tag{1.9}$$

Les vitesses de convergence dans (1.9) sont explicites, voir le Corollaire **??**. Grâce au théorème de Brown, Low (1996), nous obtenons, en conséquence, une équivalence asymptotique avec le modèle de régression

$$Y_i = \sqrt{f\Big(\frac{i}{T_n}\Big)} + \frac{1}{2\sqrt{T_n}}\xi_i, \quad \xi_i \sim \mathscr{N}(0,1), \quad i = 1, \dots, [T_n].$$

À noter que ce genre de modèle de bruit blanc gaussien s'avère être asymptotiquement équivalent à une expérience liée à l'estimation de densités non paramétriques, voir Nussbaum (1996). Pour pouvoir étendre nos résultats au cas d'une mesure de Lévy infinie, nous nous plaçons dans le cadre des modèles dominés. Pour cela, nous allons supposer l'existence d'une mesure de Lévy ν_0 qui domine toutes les mesures possibles ν et qui admet une densité g par rapport à la mesure de Lebesgue sur I. De plus, ν_0 est supposée connue et les densités (inconnues) de Lévy $f = \frac{d\nu}{d\nu_0}$, qui sont les paramètres, doivent être bornées loin de zéro et de l'infini. Les Théorèmes 4.2.5 et 4.2.6 du Chapitre 4 sont énoncés dans ce cadre général. Sous des hypothèses adéquates sur $\mathscr{F}$ (voir la Sous-section 4.2.1 pour une définition complète de $\mathscr{F}$ et la Sous-section **??** pour sa discussion), nous prouvons l'équivalence asymptotique des modèles $\mathscr{P}_n^{\nu_0}$ et $\mathscr{Q}_n^{\nu_0}$ avec un modèle de bruit blanc gaussien de la forme :

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n}}\frac{dW_t}{\sqrt{g(t)}}, \quad t \in I.$$

L'idée derrière la preuve de l'équivalence $\Delta(\mathscr{P}_n^{\nu_0}, \mathscr{Q}_n^{\nu_0}) \to 0$ est la suivante. Asymptotiquement $\mathscr{P}_n^{\nu_0}$ devrait contenir la même quantité d'information sur f que l'expérience $\mathscr{L}_m$ qui compte le nombre de sauts dont l'amplitude appartient à des intervalles $(J_i)_{i=2}^m$ formant une partition de $I \setminus [-\varepsilon_m, \varepsilon_m]$, $\varepsilon_m \to 0$. Or, les variables aléatoires $\big(\sum_{s\leq T_n} \mathbb{I}_{J_i}(\Delta X_s)\big)_{i=2}^m$ sont indépendantes de loi de Poisson de paramètres $T_n \int_{J_i} f(y)\nu_0(dy)$. On peut donc approcher $\mathscr{L}_m$ par le modèle associé à des variables aléatoires indépendantes et gaussiennes en utilisant le Théorème 4 dans Brown et al. (2004b). Grâce à une construction explicite des noyaux, on peut alors déduire l'équivalence asymptotique avec $\mathscr{W}_n^{\nu_0}$. Plusieurs difficultés sont apparues pour démontrer l'équivalence entre $\mathscr{P}_n^{\nu_0}$ et $\mathscr{L}_m$. Tout d'abord, nous avons dû prendre en compte le fait que l'intervalle I peut être de longueur infinie : pour cela, on a construit les J_i en utilisant les quantiles de ν_0 sur l'intervalle I privé d'un voisinage de 0. Plus précisément, si $x > 0$, pour tout entier strictement positif m, on a défini les intervalles $J_j := (v_{j-1}, v_j]$ où $v_1 = \varepsilon_m$ et v_j sont les quantiles de $\nu_{0|I\setminus[0,\varepsilon_m]}$, i.e.

$$\nu_0(J_j) = \frac{\nu_0\big((I \setminus [0,\varepsilon_m]) \cap \mathbb{R}_+\big)}{m-1}, \quad \forall j = 2, \dots, m. \tag{1.10}$$

On notera $\mu_n := \nu_0(J_j)$, pour tout $j = 2, \dots, m$. De manière analogue, si $x \leq 0$, nous avons défini $\mu_n^- = \frac{\nu_0\big((I \setminus [-\varepsilon_m, 0]) \cap \mathbb{R}_-\big)}{m-1}$ et $J_{-m}, \dots, J_{-2}$ tels que $\nu_0(J_{-j}) = \mu_n^-$ pour tout j.

Par un argument de statistique exhaustive, il est clair qu'il y a une équivalence entre $\mathscr{L}_m$ et le modèle $\bar{\mathscr{P}}_n^{\nu_0}$ ayant comme densité de Lévy une approximation constante par morceaux de f, que l'on notera $\bar{f}_m$ et définie de la manière suivante : si $x > 0$,

$$\bar{f}_m(x) := \begin{cases} 1 & \text{si } x \in J_1, \\ \frac{\nu(J_j)}{\nu_0(J_j)} & \text{si } x \in J_j, \quad j = 2, \dots, m, \end{cases}$$

et symétriquement pour $x < 0$ en utilisant J_j pour $j = -m, \dots, 2$. Cependant, la convergence de $\Delta(\mathscr{P}_n^{\nu_0}, \bar{\mathscr{P}}_n^{\nu_0})$ est trop lente. Pour obtenir une bonne vitesse de convergence, nous avons prouvé l'équivalence de $\bar{\mathscr{P}}_n^{\nu_0}$ avec un modèle auxiliaire $\tilde{\mathscr{P}}_n^{\nu_0}$ ayant comme densité de Lévy une adéquate approximation continue de f, notée par $\hat{f}_m$. Cette équivalence a été obtenue en construisant des noyaux markoviens.

Concernant la preuve de l'équivalence $\Delta(\mathscr{Q}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) \to 0$, nous nous sommes tout d'abord ramenés à observer les accroissements d'un processus de Poisson composé $\tilde{X}$ ayant comme densité de Lévy la fonction $\hat{f}_m$ restreinte à $I \setminus [-\varepsilon_m, \varepsilon_m]$. Ensuite, nous avons prouvé qu'observer les accroissements $(\tilde{X}_{t_i} - \tilde{X}_{t_{i-1}})_{i=1}^n$ est asymptotiquement équivalent à observer des variables aléatoires $(\varepsilon_i Y_i)_{i=1}^n$ où :

- $(\varepsilon_i)_{i=1}^n$ sont des variables aléatoires i.i.d. de loi de Bernoulli de paramètre $\alpha := \lambda_m \Delta_n e^{-\lambda_m \Delta_n}$, avec $\lambda_m = \int_{I \setminus [-\varepsilon_m, \varepsilon_m]} \hat{f}_m(y) \nu_0(dy)$,
- $(Y_i)_{i=1}^n$ sont des variables aléatoires i.i.d. concentrées sur $I \setminus [-\varepsilon_m, \varepsilon_m]$ et de densité $\hat{f}_m$.

L'étape suivante consiste à compter le nombre des réalisations de $(\varepsilon_j Y_j)_{j=1}^n$ qui tombent dans les intervalles J_i, $i = -m, \dots, -2, 2, \dots, m$. Cette expérience est associée à une famille de loi multinomiale de paramètres $n, (\gamma_{-m}, \dots, \gamma_{-2}, \gamma_0, \gamma_2, \dots, \gamma_m)$ où $\gamma_0 = 1 - \alpha$ et $\gamma_i = \alpha \int_{J_i} \hat{f}_m(y) \nu_0(dy)$. On peut alors conclure qu'il y a une équivalence asymptotique entre ce modèle et $\mathscr{W}_n^{\nu_0}$ en utilisant le résultat de Carter (2002) conjointement à certains passages techniques, notamment pour prendre en compte les problèmes à proximité de zéro.

1.2.2 Équivalence asymptotique pour des modèles à densité

Un des premiers résultats d'équivalence asymptotique au sens de Le Cam dans un contexte non paramétrique est celui de Nussbaum (1996). Dans cet article Nussbaum a montré l'équivalence asymptotique entre une expérience associée à l'observation de n variables aléatoires ayant une densité f sur $[0,1]$ et un modèle de signal avec bruit blanc :

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{n}}dW_t, \quad t \in [0,1]. \tag{1.11}$$

L'équivalence est établie en supposant que f appartienne à une classe fonctionnelle $\mathscr{F}$ définie de la manière suivante. Considérons pour $\alpha \in (0,1)$, $M > 0$, une classe de fonctions hölderiennes

$$\mathcal{H}^\alpha(M) = \{f : |f(x) - f(y)| \leq M|x-y|^\alpha\}.$$

Pour $\alpha, M, \varepsilon > 0$ donnés, nous définissons

$$\mathscr{F} \subseteq \Big\{ f \in \mathcal{H}^\alpha(M) : \int_0^1 f(x)dx = 1 \text{ et } f(x) \geq \varepsilon \ \forall x \in [0,1] \Big\}.$$

La preuve de ce résultat repose sur une technique de poissonisation et sur le couplage. Les noyaux de Markov réalisant l'équivalence asymptotique ne sont pas donnés de manière explicite. Des travaux plus récents (Carter (2002), Brown et al. (2004a)) ont réussi à obtenir le même résultat en construisant explicitement les noyaux markoviens. Le résultat de Carter (2002) est particulièrement important dans notre travail. Carter prouve l'équivalence asymptotique entre le modèle à densité et le modèle de signal avec bruit blanc (1.11) en majorant la distance de Le Cam entre des variables aléatoires de loi multinomiale et de loi normale multivariée. À partir de cette majoration, Carter a pu redémontrer la plupart des résultats de Nussbaum (1996), sous de plus fortes hypothèses de régularité sur $\mathscr{F}$: il doit supposer que $\mathscr{F}$ est une classe de densités définies sur $[0,1]$, différentiables et telles qu'il existe des constantes strictement positives ε, M, α telles que $\varepsilon \leq f \leq M$ et

$$|f'(x) - f'(y)| \leq M|x-y|^\alpha, \quad \text{pour tout } x, y \in [0,1].$$

Brièvement, on peut utiliser une borne sur la distance de Le Cam entre les variables de loi multinomiale et les variables normales multivariées pour en dériver des conclusions sur l'équivalence pour des expériences à densité. L'idée est de voir l'expérience multinomiale comme l'expérience qui compte le nombre d'observations tirées selon f qui apparaissent dans les intervalles $J_i = (\frac{i-1}{m}, \frac{i}{m}]$, $i = 1, \ldots, m$. En utilisant la racine carrée comme une transformation stabilisatrice de la variance, ces variables de loi multinomiale peuvent être asymptotiquement approchées par des variables normales avec des variances constantes. Ces variables normales, à leur tour, sont des approximations des accroissements du processus (y_t) défini par (1.11) sur les ensembles J_i.

Dans le Chapitre 5, nous nous proposons de généraliser les résultats de Nussbaum (1996) et Carter (2002). Plus précisément, l'expérience que nous considérons consiste à observer n variables aléatoires i.i.d. $(Y_i)_{i=1}^n$ définies sur un intervalle $I \subseteq \mathbb{R}$ et distribuées selon une loi P_f^g ayant une densité (par rapport à la mesure de Lebesgue sur I) $\frac{dP_f^g}{dx}(x) = f(x)g(x)$. En particulier, nous ne supposons pas que l'intervalle $I \subseteq \mathbb{R}$ soit un sous-intervalle borné de $\mathbb{R}$ comme dans la littérature existante. La fonction g est censée être connue alors que f est une fonction inconnue qui appartient à une certaine classe fonctionnelle non paramétrique $\mathscr{F}$. Formellement, le modèle statistique que nous considérons est

$$\mathscr{P}_n^g = \left(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \left(\bigotimes_{i=1}^n P_f^g : f \in \mathscr{F}\right)\right). \tag{1.12}$$

Les hypothèses sur f et g seront précisées dans la Sous-section 5.2. Pour le moment, nous nous contentons de souligner le fait que f doit être bornée loin de zéro et de l'infini, tandis que g peut être à la fois non bornée et discontinue. L'avantage par rapport aux travaux antérieurs est que ce cadre nous permet de traiter des densités de la forme $h = fg$ non nécessairement bornées ni lisses (voir la Sous-section 5.3.1 pour une discussion sur les hypothèses.)

Finalement, nous allons introduire le modèle de signal avec bruit blanc. À cet effet, nous noterons par $\mathbb{W}_f^g$ la loi induite sur $(C, \mathscr{C})$ par un processus stochastique satisfaisant l'équation différentielle stochastique :

$$dY_t = \sqrt{f(t)g(t)}dt + \frac{dW_t}{2\sqrt{n}}, \quad t \in I,$$

où $(W_t)_{t\in\mathbb{R}}$ est un mouvement Brownien sur $\mathbb{R}$ conditionné à $W_0 = 0$. Ensuite, nous définissons

$$\mathscr{W}_n^g = \big(C, \mathscr{C}, \{\mathbb{W}_f^g : f \in \mathscr{F}\}\big). \tag{1.13}$$

Le résultat principal du Chapitre 5 est alors le suivant (voir le Théorème 5.3.1 pour un énoncé précis) :

Théorème 1.2.4. *Soit I un sous-intervalle éventuellement infini de $\mathbb{R}$ et soit $\mathscr{F}$ une classe de fonctions bornées loin de zéro et de l'infini qui satisfont les hypothèses de régularité indiquées dans la Sous-section 5.2. Alors les modèles statistiques $\mathscr{P}_n^g$ et $\mathscr{W}_n^g$ définis par* (1.12) *et* (1.13), *respectivement, sont asymptotiquement équivalents :*

$$\lim_{n\to\infty} \Delta(\mathscr{P}_n^g, \mathscr{W}_n^g) = 0. \tag{1.14}$$

Dans certains cas particuliers des majorations explicites pour les vitesses de convergence dans (1.14) sont disponibles ; voir, par exemple, le Corollaire 5.3.2. La structure de la preuve suit celle de Carter (2002), on reprend notamment l'idée d'une approximation multinomiale-normale multivariée, mais certains points techniques diffèrent. Parmi ces points, l'intervalle I peut être infini et donc, en particulier, les sous-intervalles J_i qui forment une partition de I ne peuvent plus être de même longueur. Dans ce contexte, nous choisissons des intervalles J_i de longueur variable, mais également répartis selon les quantiles de ν_0, la mesure ayant une densité g par rapport à la mesure de Lebesgue. Plus précisément, l'étape clé de la preuve est l'approximation de la densité fg par une densité de la forme $\hat{f}_m g$ avec $\hat{f}_m$ définie par :

$$\hat{f}_m(x) := \begin{cases} \frac{\nu(J_1)}{\mu_n} & \text{si } x \in I \cap (-\infty, x_1^*], \\ \frac{1}{x_{j+1}^* - x_j^*}\left[\frac{\nu(J_{j+1})}{\mu_n}(x - x_j^*) + \frac{\nu(J_j)}{\mu_n}(x_{j+1}^* - x)\right] & \text{si } x \in (x_j^*, x_{j+1}^*] \quad j = 1, \dots, m-1, \\ \frac{\nu(J_m)}{\mu_n} & \text{si } x \in I \cap (x_m^*, \infty), \end{cases}$$

où

$$\mu_n := \nu_0(J_j) = \frac{\nu_0(I)}{m} \quad \text{et } x_j^* := \frac{\int_{J_j} x\nu_0(dx)}{\mu_n}, \quad \forall j = 1, \dots, m.$$

Les noyaux réalisant l'équivalence asymptotique sont explicitement construits. Remarquons aussi que le résultat présenté ci-dessus est une extension directe des résultats de Nussbaum (1996) et Carter (2002) : il suffit de choisir $g(x) = \mathbb{I}_{[0,1]}(x)$.

1.2.3 Équivalence asymptotique dans des modèles de diffusion

Dans le Chapitre 6, nous avons considéré un problème d'équivalence pour des modèles de diffusion. Plus précisément, nous nous sommes focalisés sur des diffusions unidimensionnelles de la forme :

$$dy_t = f(y_t)dt + \varepsilon\sigma(y_t)dW_t, \quad t \in [0, T], \quad y_0 = w \in \mathbb{R}, \tag{1.15}$$

où $(W_t)_{t\geq 0}$ est un $(\mathscr{A}_t)_{t\geq 0}$-mouvement Brownien standard défini sur l'espace de probabilité $(\Omega, \mathscr{A}, \mathbb{P})$.

Ce problème a intéressé plusieurs travaux présents dans la littérature. Par exemple, Delattre, Hoffmann (2002) ont établi une équivalence asymptotique entre un modèle de diffusion récurrent nul au sens de Harris et certains modèles gaussiens. D'autres résultats d'équivalence pour des diffusions ergodiques d'abord unidimensionnelles, puis multidimensionnelles, ont été considérés dans Dalalyan, Reiß (2006, 2007a). Un autre résultat sur le modèle de diffusion a été prouvé dans Genon-Catalot, Laredo, Nussbaum (2002). Dans cet article, les auteurs ont établi l'équivalence asymptotique entre une expérience mettant en jeu une diffusion transiente ayant une dérive positive et une petite variance et des expériences gaussiennes et poissoniennes.

Notre travail a consisté à comparer trois modèles statistiques : $\mathscr{P}_y^T$, $\mathscr{Q}_y^n$ et $\mathscr{Q}_Z^n$. Les deux premiers correspondent, respectivement, aux modèles associés à l'observation continue de $(y_t)_{t\in[0,T]}$ et celle discrète des variables aléatoires $(y_{t_1},\dots,y_{t_n})$, avec $t_i = Ti/n$. Le troisième modèle, $\mathscr{Q}_Z^n$, est associé à l'observation du schéma d'Euler de pas T/n, $n \in \mathbb{N}^*$, associé à $(y_t)_{t\in[0,T]}$:

$$Z_0 = w, \quad Z_i = Z_{i-1} + \frac{T}{n} f(Z_{i-1}) + \varepsilon\sqrt{\frac{T}{n}}\sigma(Z_{i-1})\xi_i, \quad i = 1,\dots,n, \tag{1.16}$$

où les ξ_i sont des variables aléatoires i.i.d. gaussiennes centrées et réduites.

Le résultat principal du Chapitre 6 montre l'équivalence asymptotique entre les expériences statistiques $\mathscr{P}_y^T$, $\mathscr{Q}_y^n$ et $\mathscr{Q}_Z^n$, en ce qui concerne l'estimation de la dérive f. Le résultat d'équivalence entre $\mathscr{Q}_y^n$ et $\mathscr{Q}_Z^n$ est particulièrement important. En effet, estimer la dérive f à partir des observations discrètes $(y_{t_1},\dots,y_{t_n})$ est un problème difficile mathématiquement, notamment car les transitions ne sont pas connues de manière explicite. En pratique, les procédures d'estimation basées sur le schéma d'Euler ont été appliquées avec succès, dans un cadre paramétrique aussi bien qu'en non paramétrique. Notre démonstration de l'équivalence au sens de Le Cam entre $\mathscr{Q}_y^n$ et $\mathscr{Q}_Z^n$ permet notamment de justifier théoriquement cette pratique.

Dans la première publication qui traite de l'équivalence statistique pour des modèles de diffusion, Milstein, Nussbaum (1998), ont prouvé l'équivalence asymptotique entre l'observation continue et discrète d'un modèle de diffusion défini par l'équation différentielle stochastique :

$$dY_t = f(Y_t)dt + \varepsilon dW_t, \quad t \in [0,1], \quad Y_0 = 0,$$

et le schéma d'Euler correspondant :

$$Z_0 = 0, \quad Z_i = Z_{i-1} + \frac{f(Z_{i-1})}{n} + \frac{\varepsilon}{\sqrt{n}}\xi_i, \quad i = 1,\dots,n,$$

où les ξ_i sont des variables aléatoires i.i.d. gaussiennes centrées et réduites. Dans ce contexte, ε est un "petit" paramètre supposé connu ne dépendant pas du temps $t \in [0,1]$. La dérive f est inconnue et telle que, pour une certaine constante $K > 0$,

$$f \in \mathscr{F}_K = \Big\{f : \mathbb{R} \to \mathbb{R} \text{ t.q. } \forall x,u \in \mathbb{R},\ |f(x) - f(u)| \le K|x-u|,\ |f(0)| \le K\Big\}.$$

Si on note respectivement $\mathscr{P}$ et $\mathscr{Z}_n$ les modèles statistiques associés à l'observation continue de $(Y_t)_{t\in[0,1]}$ et au schéma d'Euler $(Z_i)_{i=0}^n$ relatif à Y et de pas $1/n$, alors

$$\Delta(\mathscr{P},\mathscr{Z}_n) = O\Big(\sqrt{n^{-2}\varepsilon^{-2} + n^{-1}}\Big), \quad \text{quand } \varepsilon \to 0.$$

De plus, si on note $\mathscr{Q}_n$ le modèle statistique associé à l'observation discrète de $(Y_t)_{t\in[0,1]}$ avec un pas de discrétisation $\frac{1}{n}$, les auteurs ont prouvé aussi que $\Delta(\mathscr{P},\mathscr{Q}_n)=0$. Comme corollaire, ils obtiennent donc

$$\Delta(\mathscr{Q}_n,\mathscr{X}_n)=O\Big(\sqrt{n^{-2}\varepsilon^{-2}+n^{-1}}\Big),\quad \text{quand } \varepsilon\to 0.$$

Le résultat de Genon-Catalot, Laredo (2014) vise toujours à établir une équivalence entre des diffusions et un schéma d'Euler. Plus en détail, les auteurs considèrent une diffusion (ξ_t) donnée par :

$$d\xi_t=b(\xi_t)dt+\sigma(\xi_t)dW_t,\quad \xi_0=\eta,$$

où (W_t) est un mouvement Brownien standard défini sur un espace de probabilité filtré $(\Omega,\mathscr{A},(\mathscr{A}_t)_{t\geq 0},\mathbb{P})$, η est une variable aléatoire réelle, $\mathscr{A}_0$-mesurable et $b(\cdot)$, $\sigma(\cdot)$ sont des fonctions définies sur $\mathbb{R}$. Le coefficient de diffusion $\sigma(\cdot)$ est une fonction dans $C^2(\mathbb{R})$, supposée connue, qui satisfait les conditions suivantes :

$$\forall x\in\mathbb{R},\ 0<\sigma_0^2\leq\sigma^2(x)\leq\sigma_1^2,\quad |\sigma'(x)|+|\sigma''(x)|\leq K_\sigma.$$

Le processus (ξ_t) est observé soit d'une manière continue, soit d'une manière discrète tout au long de l'intervalle temporel $[0,T]$. Les observations discrètes ont la forme ξ_{t_i}, $t_i=ih$, $i\leq n$ avec $T=nh$. Le résultat principal est l'équivalence asymptotique, en ce qui concerne la dérive b, entre l'observation discrète et continue de $(\xi_t)_{t\in[0,T]}$ et le schéma d'Euler correspondant :

$$Z_0=\eta,\quad Z_i=Z_{i-1}+hb(Z_{i-1})+\sqrt{h}\sigma(Z_{i-1})\varepsilon_i,$$

où, pour $i\geq 1$, $t_i=ih$ et $\varepsilon_i=(W_{t_i}-W_{t_{i-1}})/\sqrt{h}$. Les équivalences sont établies sous l'hypothèse que $h=h_n$ et $nh_n^2=T^2/n$ tendent vers 0 lorsque n tend vers l'infini. Cela comprend à la fois les cas $T=nh_n$ borné et $T\to\infty$.

Cependant, pour pouvoir prendre en compte un intervalle de temps pouvant être infini, les auteurs sont obligés de mettre des hypothèses plus fortes sur l'espace de paramètre $\mathscr{F}$ par rapport aux hypothèses de Milstein, Nussbaum (1998). Plus précisément, les équivalences établies dans Genon-Catalot, Laredo (2014) sont valables sous l'hypothèse que la dérive $b(\cdot)$ appartienne à une classe fonctionnelle $\mathscr{F}_K$ pour une certaine constante positive K :

$$b(\cdot)\in\mathscr{F}_K=\Big\{b(\cdot)\in C^1(\mathbb{R})\text{ et pour tout } x\in\mathbb{R},\ |b(x)|+|b'(x)|\leq K\Big\}.$$

Si on applique le résultat d'équivalence de Genon-Catalot, Laredo (2014) au cas de la petite variance, c'est-à-dire, si on prend un coefficient de diffusion de la forme $\varepsilon\sigma(\cdot)$ avec ε petit et un horizon temporel fixe T, on retrouve la condition $\varepsilon n\to\infty$ (comme dans le Milstein, Nussbaum (1998)). À cela s'ajoute une condition technique qui est $\varepsilon^4 n \to \infty$. Plus précisément, une majoration de la vitesse de convergence entre les observations discrètes (associées au modèle $\mathscr{Q}_n$) et le schéma d'Euler (associé au modèle $\mathscr{X}_n$) est donnée par :

$$\Delta(\mathscr{Q}_n, \mathscr{X}_n) = O\Big(\sqrt{n^{-2}\varepsilon^{-2} + n^{-1} + n^{-1}\varepsilon^{-4}}\Big).$$

Notons que dans le Chapitre 6 nous avons trouvé la borne

$$\Delta(\mathscr{Q}_n, \mathscr{X}_n) = O\Big((n^{-1} + \varepsilon)^{1/4} + n^{-1}\varepsilon^{-1}\Big).$$

Pour obtenir le résultat présenté dans le Chapitre 6, nous nous sommes concentrés uniquement sur le cas de la petite variance avec un coefficient de diffusion $\sigma(\cdot)$ non constant. Nous avons obtenu un résultat d'équivalence entre un processus de diffusion et le schéma d'Euler correspondant avec les hypothèses de Milstein, Nussbaum (1998), qui sont plus faibles que celles de Genon-Catalot, Laredo (2014). Pour le démontrer nous nous sommes éloignés des techniques de preuve utilisées dans ce dernier article. Nous avons utilisé le fait que le processus $(y_t)_{t\in[0,T]}$ défini par (1.15) converge vers une solution déterministe dans le cas de la petite variance. Avant de rentrer dans le détail des techniques utilisées dans la preuve, commençons par formaliser les expériences statistiques ainsi que l'espace des paramètres considérés.

Concernant les hypothèses sur f, nous devons supposer les conditions classiques pour l'existence et l'unicité d'une solution forte de l'équation différentielle stochastique (1.15) (voir Øksendal (1985), Théorème 5.5). Nous allons donc travailler avec des espaces de paramètres inclus dans l'ensemble $\mathscr{F}_M$ de toutes les fonctions f définies sur $\mathbb{R}$ et qui satisfont :

$$|f(0)| \leq M \text{ and } |f(z) - f(y)| \leq M|z-y|, \quad \forall z, y \in \mathbb{R}.$$

Le coefficient de diffusion $\varepsilon\sigma(\cdot)$, avec $0 < \varepsilon < 1$, est supposé connu et il doit satisfaire les hypothèses suivantes :

(H1) $\sigma(\cdot)$ est une fonction réelle K-lipschitzienne bornée loin de zéro et de l'infini, c'est-à-dire, il existe des constantes strictement positives σ_0, σ_1, K telles que :

$$\sigma_0^2 \leq \sigma^2(y) \leq \sigma_1^2 \text{ et } |\sigma(z) - \sigma(y)| \leq K|z-y|, \quad \forall z, y \in \mathbb{R}.$$

Lorsque $(y_t)_{t\in[0,T]}$ est observé de manière discrète, nous supposerons également que :

(H2) $\sigma(\cdot)$ est une fonction différentiable sur $\mathbb{R}$ avec une dérivée K-lipschitzienne, c'est-à-dire :

$$|\sigma'(z) - \sigma'(y)| \leq K|z-y| \quad \forall z, y \in \mathbb{R}.$$

De plus, la distribution induite sur $(C, \mathscr{C}_T)$ par la loi du processus $(y_t)_{t\in[0,T]}$ solution de (1.15) sera notée $P_f^{n,y}$ et la distribution définie sur $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ par la loi de $(y_{t_1}, \ldots, y_{t_n})$, $t_i = Ti/n$ sera notée $Q_f^{n,y}$. Nous appellerons $\mathscr{P}_y^T$ l'expérience associée à l'observation continue de $(y_t)_{y\in[0,T]}$ et $\mathscr{Q}_y^n$ l'expérience associée à l'observation discrète basée sur les variables y_{t_i}, $i = 1, \ldots, n$:

$$\mathscr{P}_y^T = \big(C, \mathscr{C}_T, \{P_f^y, f \in \mathscr{F}\}\big),$$
$$\mathscr{Q}_y^n = \big(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \{Q_f^{n,y}, f \in \mathscr{F}\}\big).$$

Finalement, nous noterons $Q_f^{n,Z}$ la distribution du vecteur $(Z_i, i = 1, \ldots, n)$ défini par l'équation (1.16) et $\mathscr{Q}_Z^n$ le modèle statistique associé à la famille de lois $(Q_f^{n,Z}, f \in \mathscr{F})$:

$$\mathscr{Q}_Z^n = \big(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \{Q_f^{n,Z}, f \in \mathscr{F}\}\big).$$

Nos résultats principaux sont les suivants.

Théorème 1.2.5. *On note $\mathscr{F}$ l'espace des paramètres. Soit $M > 0$ tel que $\mathscr{F} \subseteq \mathscr{F}_M$ et $\sigma(\cdot)$ vérifie l'hypothèse (H1) avec $K = M$. Alors, si $\varepsilon n \to \infty$ lorsque $n \to \infty$ et $\varepsilon \to 0$, les expériences $\mathscr{P}_y^T$ et $\mathscr{Q}_Z^n$ sont asymptotiquement équivalentes. Plus précisément, une borne supérieure pour la vitesse de convergence de la distance de Le Cam est donnée par :*

$$\Delta\big(\mathscr{P}_y^T, \mathscr{Q}_Z^n\big) = O\Big(\frac{1}{\varepsilon n} + (n^{-1} + \varepsilon)^{1/4}\Big).$$

Théorème 1.2.6. *On note $\mathscr{F}$ l'espace des paramètres. Soit $M > 0$ tel que $\mathscr{F} \subseteq \mathscr{F}_M$ et $\sigma(\cdot)$ vérifie les hypothèses (H1) et (H2) avec $K = M$. De plus, supposons que $\sigma(\cdot)$ et $\mathscr{F}$ soient telles que $f/\sigma(\cdot)$ est une fonction L-lipschitzienne avec une constante L uniforme pour toute $f \in \mathscr{F}$. Alors, pour tout ε, éventuellement fixé, les valeurs échantillonnées $y_{t_1}, \ldots, y_{t_n}$ constituent une statistique asymptotiquement exhaustive pour l'expérience $\mathscr{P}_y^T$.*

Corollaire 1.2.7. *Sous les mêmes hypothèses que dans le Théorème 1.2.6, le modèle statistique associé aux valeurs échantillonnées $y_{t_1}, \ldots, y_{t_n}$ est asymptotiquement équivalent au modèle $\mathscr{Q}_Z^n$, lorsque n tend vers l'infini. On retrouve la même borne supérieure pour la vitesse de convergence de la distance de Le Cam énoncée dans le Théorème 1.2.5.*

Comme souligné au début de cette section, le résultat qui a le plus d'intérêt pratique est celui énoncé dans le Corollaire 1.2.7. Il est obtenu comme conséquence de l'équivalence asymptotique entre $\mathscr{P}_y^T$ et $\mathscr{Q}_Z^n$, d'un côté et entre $\mathscr{P}_y^T$ et $\mathscr{Q}_y^n$ de l'autre. Admettre un coefficient de diffusion non constant a engendré plusieurs difficultés dans la preuve. Dans ce qui suit, nous essayons d'expliquer d'où elles proviennent. Pour cela, rappelons brièvement la preuve du résultat de Milstein, Nussbaum (1998) :

- On peut montrer aisément que le modèle associé au schéma d'Euler est équivalent à un modèle de diffusion ayant des coefficients qui sont des "discrétisations" de f. Notons $(\bar{Y}_t)_{t\in[0,T]}$ ce modèle de diffusion.

- On compare les modèles associés aux observations continues de $(Y_t)_{t\in[0,T]}$ et $(\bar{Y}_t)_{t\in[0,T]}$ grâce au fait que le coefficient de diffusion est constant, et une application du théorème de Girsanov. Ceci permet de contrôler la distance en variation totale entre les lois induites par Y et $\bar{Y}$ sur $(C, \mathscr{C}_T)$.

- Les deux points précédents permettent de conclure à l'équivalence asymptotique entre $\mathscr{P}_y^T$ et $\mathscr{Q}_Z^n$. Pour déduire l'équivalence entre $\mathscr{P}_y^T$ et $\mathscr{Q}_y^n$, il suffit d'observer que, dans le cas d'un coefficient de diffusion constant, la statistique $S : \omega \mapsto (\omega_{t_1}, \dots, \omega_{t_n})$ est exhaustive pour la famille des lois $(P_f^{n,y} : f \in \mathscr{F})$.

Cependant, ce schéma de preuve ne peut pas être généralisé au cas d'un coefficient de diffusion non constant. En effet, avec un coefficient de diffusion non constant, nous avons toujours l'équivalence entre le schéma d'Euler et un processus de diffusion $\bar{y}$ qui satisfait l'équation différentielle stochastique :

$$d\bar{y}_t = \bar{f}_n(t, \bar{y})dt + \varepsilon\bar{\sigma}_n(t, \bar{y})dW_t, \quad \bar{y}_0 = w, \quad t \in [0, T],$$

où, pour toute fonction ω dans C, et en utilisant la notation $t_i = Ti/n$, nous définissons :

$$\bar{f}_n(t, \omega) = \sum_{i=1}^{n-1} f\big(\omega(t_i)\big)\mathbb{I}_{[t_i, t_{i+1})}(t), \quad \bar{\sigma}_n(t, \omega) = \sum_{i=1}^{n-1} \sigma\big(\omega(t_i)\big)\mathbb{I}_{[t_i, t_{i+1})}(t), \quad \forall t \in [0, T].$$

Étant donné que le coefficient de diffusion du processus y est différent de celui de $\bar{y}$, la distance en variation totale entre les lois de y et de $\bar{y}$ est toujours égale à 1. Donc, pour pouvoir comparer les expériences associées aux processus y et $\bar{y}$, la construction d'un noyau markovien s'impose. Construire directement un tel noyau s'est révélé infaisable dans la pratique et nous avons décidé de contourner le problème en introduisant des expériences changées en temps, comme cela a déjà été proposé par Genon-Catalot, Laredo (2014). Cela permet alors de travailler avec des diffusions en petite variance ayant un coefficient de diffusion identiquement égal à ε, pour lesquelles nous sommes donc capables de comparer la distance en variation totale. Le prix à payer est celui de devoir ainsi gérer des expériences observées jusqu'à différents instants aléatoires : c'est exactement dans le contrôle de ces différents temps d'arrêt que la présence de la petite variance joue un rôle fondamental.

1.2.4 Distance L_1 entre processus additifs

Donner des majorations pour la distance L_1 entre des lois définies sur l'espace de Skorokhod est un problème classique. Par distance L_1 entre deux mesures σ-finies μ_1 et μ_2 définies sur un espace mesurable $(E, \mathscr{E})$, nous entendons :

$$L_1(\mu_1, \mu_2) = 2 \sup_{A \in \mathscr{E}} \big|\mu_1(A) - \mu_2(A)\big|,$$

c'est-à-dire que la distance L_1 est définie comme étant le double de la distance en variation totale. Si μ_1 est absolument continue par rapport à μ_2, nous avons aussi la définition :

$$L_1(\mu_1, \mu_2) = \int_E \left|\frac{d\mu_1}{d\mu_2} - 1\right| d\mu_2.$$

Le contrôle des distances entre deux lois apparaît dans plusieurs domaines tels que la statistique bayésienne, le contrôle des vitesses de convergence des chaînes de Markov ou les algorithmes de Monte Carlo.

Dans le Chapitre 7, nous proposons une majoration de la distance L_1 entre des processus additifs. Plus précisément, nous avons comparé deux processus additifs X_i ayant comme caractéristique locale (f_i, σ_i^2, ν_i), $i = 1, 2$. Nous avons supposé $\sigma_1 = \sigma_2$ car, sinon, la distance L_1 entre les lois de X_1 et X_2 vaut constamment 2 (voir, par exemple, Jacod, Shiryaev (2003) ; Newman (1972)). De plus, définissons γ^{ν_j} et ξ^2 tels que :

$$\gamma^{\nu_j} = \int_{|y| \leq 1} y \nu_j(dy), \quad \xi^2 = \int_0^T \frac{(f_2(r) - f_1(r) - (\gamma^{\nu_2} - \gamma^{\nu_1}))^2}{\sigma^2(r)} dr \quad j = 1, 2.$$

Le résultat principal du Chapitre 7 est alors le suivant.

Théorème 1.2.8. *Soient X_i, $i = 1, 2$, deux processus additifs ayant comme caractéristiques locales (f_i, σ_i^2, ν_i), $i = 1, 2$. Supposons que ν_1 et ν_2 soient deux mesures de Lévy avec ν_1 absolument continue par rapport à ν_2 et telles que $L_1(\nu_1, \nu_2) < \infty$. Notons ϕ la fonction de répartition d'une variable aléatoire normale $\mathscr{N}(0, 1)$ et supposons $0 < T < \infty$. Si $\sigma^2 > 0$, alors :*

$$L_1\Big(P_T^{(f_1, \sigma^2, \nu_1)}, P_T^{(f_2, \sigma^2, \nu_2)}\Big) \leq 2 \sinh\Big(T L_1(\nu_1, \nu_2)\Big) + 2\bigg[1 - 2\phi\Big(-\frac{\xi}{2}\Big)\bigg].$$

Si $\sigma^2 = 0$ et $f_1 - f_2 \equiv \gamma^{\nu_1} - \gamma^{\nu_2}$, alors :

$$L_1\Big(P_T^{(f_1, 0, \nu_1)}, P_T^{(f_2, 0, \nu_2)}\Big) \leq 2 \sinh\Big(T L_1(\nu_1, \nu_2)\Big).$$

Remarquons que, dans le cas où $\nu_1 = \nu_2 = 0$, c'est-à-dire quand il n'y a pas de sauts, la majoration qui apparaît dans l'énoncé du Théorème 1.2.8 est plutôt une égalité. En effet, une formule pour la distance L_1 entre processus gaussiens observés jusqu'à l'instant $0 < T < \infty$, est :

$$L_1\big(P_T^{(f_1,\sigma^2,0)}, P_T^{(f_2,\sigma^2,0)}\big) = 2\bigg(1 - 2\phi\bigg(-\frac{1}{2}\sqrt{\int_0^T \frac{(f_1(t)-f_2(t))^2}{\sigma^2(t)}dt}\bigg)\bigg),$$

quand le terme de droite est bien défini (voir, par exemple, Brown, Low (1996)).

Nous proposons une preuve basée uniquement sur des résultats classiques de processus additifs, en s'appuyant notamment sur des transformations de type Esscher et la formule de Cameron-Martin. Il s'avère que cette méthode permet d'obtenir un meilleur contrôle, par rapport à la littérature existante (voir le début du Chapitre 7 pour une comparaison entre le Théorème 1.2.8 ci-dessus énoncé et le résultat de Memin, Shiryayev (1985)).

1.3 Liste des travaux de l'auteur

Ci-dessous nous donnons une liste des travaux ayant contribué à la rédaction de la thèse. Nous avons mis entre parenthèses le chapitre de la thèse qui correspond à l'article cité.

Publications dans des revues internationales avec comité de lecture

1. Mariucci E. (2015). 'Asymptotic equivalence for inhomogeneous jump diffusion processes and white noise'. Accepté pour publication dans ESAIM : Probability and Statistics (Chapitre 3).

2. Mariucci E. (2015). 'Asymptotic equivalence of discretely observed diffusion processes and their Euler scheme : small variance case'. Accepté pour publication dans Statistical Inference for Stochastic Processes (Chapitre 6).

3. Étoré P., Mariucci E. (2014). 'L_1-distance for additive processes with time-homogeneous Lévy measures'. Electronic Communications in Probabability 19, no. 57 (Chapitre 7).

Articles soumis à des revues internationales avec comité de lecture

1. Mariucci E. 'Asymptotic equivalence of Lévy density estimation and Gaussian white noise' (Chapitre 4).

2. Mariucci E. 'Asymptotic equivalence for density estimation and Gaussian white noise : An extension' (Chapitre 5).

Chapter 2

Statistical experiments and their comparison

Résumé Le Chapitre 2 est dédié à la notion de la distance de Le Cam. Nous en rappelons la définition ainsi que les propriétés principales. Nous passons en revue des outils classiques pour majorer une telle distance avant de présenter des exemples.

Mot clés: Expériences statistiques, distance de Le Cam, déficience, transitions.

Abstract In Chapter 2 we recall the main concepts on the Le Cam theory of statistical experiments, especially the notion of Le Cam distance and its properties. We also review classical tools for bounding such a distance before presenting some examples.

Keywords Statistical experiments, Le Cam distance, deficiency, randomizations.

2.1 Introduction

The theory of *Mathematical Statistics* is based on the notion of *statistical models*, also called *statistical experiments*. A statistical model, as in its original formulation due to Blackwell (1951) is a triple

$$\mathscr{P} = (\Omega, \mathscr{T}, (P_\theta)_{\theta\in\Theta}),$$

where $(\Omega, \mathscr{T})$ is a sample space, Θ is a set called the *parameter set* and $(P_\theta)_{\theta\in\Theta}$ is a family of probability measures on $(\Omega, \mathscr{T})$. This definition is a mathematical abstraction intended to represent a concrete experiment; consider for example the following situation taken from the book of Le Cam, Yang (2000). A physicist decides to estimate the half life of Carbon 14, C^{14}. He supposes that the life of a C^{14} atom has an exponential distribution with parameter θ and, in order to develop his investigation, he takes a sample of n atoms of C^{14}. The physicist fixes in advance the duration of the experiment, say 2 hours, and then he counts the number of disintegrations. Formally, this leads to the definition of the statistical model $\mathscr{P}_1 = (\mathbb{N}, \mathcal{P}(\mathbb{N}), (P_\theta)_{\theta \in (0,\infty)})$ where P_θ represents the law of the random variable X counting the number of disintegrations observed in 2 hours. This is not the only way to proceed if we want to estimate the half life of Carbon 14. Indeed, the physicist could choose to consider the first random time Y after which a fixed number of disintegrations, say 10^6, have occurred. In this case he will represent the experiment via the statistical model $\mathscr{P}_2 = (\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+), (Q_\theta)_{\theta \in (0,\infty)})$ where Q_θ is the law of the random variable Y. A natural question is then how much “statistical information” the considered experiments contain or, more precisely, when the experiment $\mathscr{P}_1$ will be more informative than $\mathscr{P}_2$ and conversely.

The quest for comparison of statistical experiments was initiated by the paper of Bohnenblust, Shapley, Sherman (1949) followed by the papers of Blackwell (1951, 1953) where the following definition was introduced: “$\mathscr{P}_1$ is more informative than $\mathscr{P}_2$” if for any bounded loss function W and any decision procedure ρ_2 in the experiment $\mathscr{P}_2$ there exists a decision procedure ρ_1 in the experiment $\mathscr{P}_1$ such that

$$R(\mathscr{P}_1, \rho_1, W, \theta) \le R(\mathscr{P}_2, \rho_2, W, \theta), \quad \forall \theta \in \Theta.$$

Here we denote by $R(\mathscr{P}_1, \rho_1, W, \theta)$ and $R(\mathscr{P}_2, \rho_2, W, \theta)$ the *statistical risk* for the experiments $\mathscr{P}_1$ and $\mathscr{P}_2$, respectively.

However, this can lead to two models being non-comparable. This issue was solved by Le Cam who introduced the notion of deficiency $\delta(\mathscr{P}_1, \mathscr{P}_2)$. We will give a precise definition in the forthcoming sections. Here, we only remark two interesting properties:

- $\delta(\mathscr{P}_1, \mathscr{P}_2)$ is a well defined non-negative real number for every two given statistical models $\mathscr{P}_1$ and $\mathscr{P}_2$ sharing the same parameter space.

- For every loss function W with $0 \le W \le 1$ and every decision procedure ρ_2 available on Θ using $\mathscr{P}_2$, there exists a decision procedure ρ_1 in $\mathscr{P}_1$ such that for all $\theta \in \Theta$,

$$R(\mathscr{P}_1, \rho_1, W, \theta) \le R(\mathscr{P}_2, \rho_2, W, \theta) + \delta(\mathscr{P}_1, \mathscr{P}_2).$$

This solves the issue mentioned above: It could be that both $\delta(\mathscr{P}_1, \mathscr{P}_2)$ and $\delta(\mathscr{P}_2, \mathscr{P}_1)$ are strictly positive, in which case they will not be comparable according to the first definition; nevertheless, we can still say “how much information” we lose when passing from one model to the other one. Le Cam's theory has found applications in several problem in statistical decision theory and it has been developed, for example, for nonparametric regression, nonparametric density estimation problems, generalized linear models, diffusion models, Lévy models, spectral density estimation problem. Another new research direction that has been explored involves quantum statistical experiments (see, e.g. Buscemi (2012)). In this survey paper, after giving equivalent definitions of the Le Cam distance and techniques to bound it in Sections 2–5, we will review some particularly interesting equivalence results among the recent literature in Section 6.

2.2 Deficiency and Le Cam distance

In this section we will present the definition of the Le Cam distance in its most general form by using "transitions", which are a generalization of the concept of Markov kernels.

2.2.1 General definition

Definition 2.2.1 (Lattice)**.** An ordered set $(E, \leq)$ is called a *lattice* if every finite set of elements of E possesses both a supremum and an infimum. A lattice is called *complete* if all bounded subsets possess suprema and infima.

Definition 2.2.2 (Vector lattice)**.** E is called a *vector lattice* or *Riesz space* if $(E, \leq)$ is a real vector space with a lattice structure $\leq$ compatible with its vector structure, i.e.:

- if $x \leq y$ then $\forall z \in E$ $x + z \leq y + z$,
- if $x \leq y$ then $\lambda x \leq \lambda y$ $\forall \lambda \in \mathbb{R}^+$.

Notation 2.2.3. $x^+ = \sup(x, 0),\ x^- = -\inf(x, 0),\ |x| = x^+ + x^-$.

Definition 2.2.4 (Compatible norm)**.** A norm $\|\cdot\|$ on the vector lattice $(E, \leq)$ is *compatible* with the order if $|x| \leq |y|$ implies $\|x\| \leq \|y\|$.

Definition 2.2.5 (Banach lattice)**.** A *Banach lattice* $(E, \|\cdot\|)$ is a vector lattice with a compatible norm $\|\cdot\|$ for which E is complete.

Definition 2.2.6 (L-space)**.** The Banach lattice $(E, \|\cdot\|)$ is called an *L-space* if its norm satisfies: $x, y \geq 0 \implies \|x + y\| = \|x\| + \|y\|$.

Definition 2.2.7 (Positive elements and unity)**.** Let M be the dual of L. We will say that an element $u \in M$ is *positive* if $\langle u, \phi^+ \rangle \geq 0$ for every $\phi \in L$. For this order, M is a vector lattice. It has a *unit* I defined by $I\phi = \|\phi^+\| - \|\phi^-\|$.

Definition 2.2.8 (Uniform lattice)**.** Let A be a set and let Γ be a set of bounded real valued functions defined on A. Such a set will be called a *uniform lattice* if it satisfies the following conditions:

(i) Γ is a vector space and a vector lattice for the operations carried out point wise;

(ii) the function I, identically equal to unity on A is an element of Γ;

(iii) for the norm defined by $\|\gamma\| = \inf\{\alpha : |\gamma| \leq \alpha I\}$ the space Γ is complete.

Remark 2.2.9. For this norm the space Γ has a dual Γ^* which is an L-space for the order and the norm induced by Γ.

Definition 2.2.10 (Transitions)**.** Let $(L', \|\cdot\|_{L'})$ and $(L'', \|\cdot\|_{L''})$ be two L-spaces. A *transition* (or *randomization*) $T : L' \to L''$ is a positive linear map such that $\|T\mu^+\|_{L''} = \|\mu^+\|_{L'}$ for all $\mu \in L'$.

Definition 2.2.11. Let Θ be a given set. An *experiment* or a *statistical model* $\mathscr{P}$ indexed by the set Θ is a map $\theta \mapsto P_\theta$ from Θ to a certain (and abstract) L-space $(L, \|\cdot\|)$ subject to the restriction that each P_θ is positive and that $\|P_\theta\| = 1$, for all $\theta \in \Theta$.

Notation 2.2.12. *To each experiment $\mathscr{P}$ one can associate the smallest L-space which contains all the P_θ. This space is called the L-space associated with $\mathscr{P}$ and it is denoted by $L(\mathscr{P})$.*

Definition 2.2.13. Let $\mathscr{P}_i : \theta \to P_{i,\theta}$, $i = 1, 2$, be two experiments indexed by Θ. The *deficiency* $\delta(\mathscr{P}_1, \mathscr{P}_2)$ of $\mathscr{P}_1$ with respect to $\mathscr{P}_2$ is the number

$$\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_T \sup_{\theta \in \Theta} \|TP_{1,\theta} - P_{2,\theta}\|,$$

for an infimum taken over all transitions T from the L-space $L(\mathscr{P}_1)$ to $L(\mathscr{P}_2)$.

When $\delta(\mathscr{P}_1, \mathscr{P}_2) = 0$ (i.e. if the experiment $\mathscr{P}_2$ can be reconstructed from the experiment $\mathscr{P}_1$ by a randomization) it is natural to say that $\mathscr{P}_2$ is *less informative* than $\mathscr{P}_1$, or that $\mathscr{P}_1$ is *better* than $\mathscr{P}_2$, or that $\mathscr{P}_1$ is *more informative* than $\mathscr{P}_2$.

Definition 2.2.14. The *Le Cam distance* or Δ-*distance* between $\mathscr{P}_1$ and $\mathscr{P}_2$ is defined as

$$\Delta(\mathscr{P}_1, \mathscr{P}_2) = \max(\delta(\mathscr{P}_1, \mathscr{P}_2), \delta(\mathscr{P}_2, \mathscr{P}_1)).$$

If $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$ we shall call that $\mathscr{P}_1$ and $\mathscr{P}_2$ are *equivalent* or of the *same type.*

The Δ-distance is a pseudo metric on the space of all statistical models. It satisfies the triangular inequality $\Delta(\mathscr{P}_1,\mathscr{P}_3) \leq \Delta(\mathscr{P}_1,\mathscr{P}_2) + \Delta(\mathscr{P}_2,\mathscr{P}_3)$, but the equality $\Delta(\mathscr{P}_1,\mathscr{P}_2) = 0$ does not imply that $\mathscr{P}_1$ and $\mathscr{P}_2$ actually coincide. For a given set Θ the experiment types indexed by Θ form a set $E(\Theta)$ which can be shown to be complete for the pseudo-metric Δ. The set $E(\Theta)$ always has a weakest element. This is the trivial experiment where all the P_θ are the same. It also has a strongest element called the *perfect experiment*. It can be obtained by assigning to each θ the Dirac measure δ_θ.

2.3 Decision theory

In Section 2.2 we have seen that, given any two statistical models sharing the same parameter space, it is possible to compare them quantifying the cost needed to reconstruct one model from the other, via randomizations. But, as we said in the introduction, a way to compare statistical models that seems just as natural is to compare the respective risk functions. The aim of this section is to show how the definition of the deficiency has a clear interpretation in terms of statistical decision theory. To that aim, we will start by recalling the standard framework:

- A *statistical model*, which is just an indexed set $\mathscr{P} = (P_\theta : \theta \in \Theta)$ of probability measures all defined on the same measurable space $(\mathscr{X},\mathscr{T})$, for some set $\mathscr{X}$ equipped with a σ-field $\mathscr{T}$. The elements of Θ are sometimes called the *states of Nature*.

- A space Z of possible actions or decisions that the statistician can take after observing $x \in \mathscr{X}$. For example, in estimation problems we can take $Z = \Theta$. To make sense of the notion of integral on Z we need it to be equipped with a σ-field $\mathcal{Z}$.

- A loss function $W : \Theta \times Z \mapsto (-\infty, \infty]$, with the interpretation that action $z \in Z$ incurs a loss $W(\theta, z)$ when θ is the true state of Nature.

Moreover, the standard interpretation of risk is as follows. The statistician observes a value $x \in \mathscr{X}$ obtained from a probability measure P_θ. He does not know the value of θ and must take a decision $z \in Z$. He does so by choosing a probability measure $\rho(x,\cdot)$ on Z and picking a point in Z at random according to $\rho(x,\cdot)$. If he has chosen z when the true distribution of x is P_θ, he suffers a loss $W(\theta,z)$. His average loss when x is observed is then $\int W(\theta,z)\rho(x,dz)$. His all over average loss when x is picked according to P_θ is the integral $\int \big(\int W(\theta,z)\rho(x,dz)\big)P_\theta(dx)$.

In order to highlight the link between the Le Cam distance and the decision theory, we come back to Le Cam's formalism, as in Section 2.2.

Let Γ be an uniform lattice on a set Z. Consider an experiment $\mathscr{P}$ indexed by a set Θ and two additional objects:

(i) a set Z of "possible decisions";

(ii) a loss function W which maps $\Theta \times Z$ into the set $(-\infty, \infty]$.

It will always be assumed that for each $\theta \in \Theta$ one has

$$\inf_{z \in Z} W(\theta, z) > -\infty.$$

In addition we shall consider only decision problems which are *regular* in the sense that one is given a certain uniform lattice Γ on the set Z.

Definition 2.3.1 (Decision procedure)**.** A *decision procedure* ρ is a transition from $L(\mathscr{P})$ to the dual Γ^* of Γ.

Definition 2.3.2 (Risks)**.** Each decision rule has a *risk function*, the function of θ defined by the expected losses for the (non randomized) rule:

$$R(\theta, \rho) = \int W(\theta, \rho(x)) P_\theta(dx).$$

More generally, a randomized decision rule ρ has its associated risk function

$$R(\mathscr{P}, \rho, W, \theta) = \int \int W(\theta, z) \rho(x, dz) P_\theta(dx).$$

The main idea of statistical decision theory is: Decision rules should be compared using only their risk functions. If $R(\mathscr{P}, \rho, W, \theta) \leq R(\mathscr{P}, \rho', W, \theta)$ for all $\theta \in \Theta$, with strict inequality for at least one θ, then ρ' is inferior to ρ. We can also go further and summarize the virtues of a rule by means of its *minimax risk*

$$R_{mmax}(\mathscr{P}, \rho, W, \Theta) = \sup_{\theta \in \Theta} R(\mathscr{P}, \rho, W, \theta)$$

or its *Bayes risk* for some prior probability μ on Θ:

$$R(\mu, \mathscr{P}, \rho, W, \Theta) = \int R(\mathscr{P}, \rho, W, \theta) \mu(d\theta).$$

Of course Θ should be equipped with its own σ-field $\mathscr{A}$ and the loss function should be $\mathscr{A} \otimes \mathcal{Z}$-measurable.

When we wrote that the deficiency is linked to the comparison of risk functions we were deliberately evasive. From the above definitions it is clear that a Bayesian statistician might prefer a notion of deficiency in terms of Bayes risk while another one might prefer to compare maximum risks. One might also try to measure the difficulty of constructing approximations to $\mathscr{T}_2$ on the basis of $\mathscr{T}_1$ (here we are supposing that one is dealing with statistical models $\mathscr{P}_i = (\Omega_i, \mathscr{T}_i, (P_{i,\theta})_{\theta \in \Theta})$). The surprising thing is that the notion of deficiency described in Section 2.2 could have just as well been introduced from any of these point of view. This will be mathematically formalized in Theorem 2.3.6. In order to be able to state it we need to introduce further notations and definitions.

Definition 2.3.3 (Restriction of an experiment)**.** If $S \subset \Theta$ and $\mathscr{P} = (P_\theta : \theta \in \Theta)$ the map $\theta \mapsto P_\theta$ restricted to $\theta \in S$ defines an experiment indexed by S. It will be denoted by $\mathscr{P}_S$. For two such experiments $\mathscr{P}_1$ and $\mathscr{P}_2$ the deficiency $\delta(\mathscr{P}_{1,S}, \mathscr{P}_{2,S})$ will also be called the deficiency of $\mathscr{P}_1$ with respect to $\mathscr{P}_2$ on S.

It is also convenient to introduce a particular class $\mathscr{C}$ of decision spaces as follows. The class $\mathscr{C}$ consists of those decision spaces (A, Γ, W) where

(i) A is a finite set;

(ii) Γ is the space of continuous functions on A;

(iii) the loss function W is such that $0 \leq W(\theta, z) \leq 1$ for all $\theta \in \Theta$ and $z \in A$.

Finally, let us denote by Γ' the the topological dual of Γ,

$$\Gamma' = \{T : \Gamma \to \mathbb{R} \text{ such that } T \text{ is linear and continuous}\}$$

and by $\sigma(M, L)$ the weakest topology on M that makes all the element of L continuous.

Definition 2.3.4 (Characteristic envelope)**.** Let $\mathscr{R}$ be the space of all functions $f \colon \Theta \to (-\infty, \infty]$ such that $f \geq g$ for some $g : \Theta \to R(\mathscr{P}, \rho, W, \Theta)$. Let $K(\Theta)$ be the cone of all positive measures that have finite support on Θ. For each $\mu \in K(\Theta)$ let

$$\chi(\mu, \mathscr{P}, W) = \inf \Big\{ \int f d\mu; f \in \mathscr{R} \Big\}.$$

Definition 2.3.5. Let L be an L-space with dual M and let Γ be an uniform lattice. Let H be a linear subspace of M which is a sublattice in the sense that $u \in H$ implies $|u| \in H$. Assume also that H is a uniform lattice on a set A. A transition T from L to Γ^* will be called (Γ, H)-*continuous* if for every $\gamma \in \Gamma$ the image $\gamma T : L \to \mathbb{R}$ belongs to H.

Theorem 2.3.6. *[Le Cam (1986)] Let $\mathscr{P}_1 = (P_\theta : \theta \in \Theta)$ and $\mathscr{P}_2 = (Q_\theta : \theta \in \Theta)$ be two experiments indexed by the set Θ. Let $\varepsilon \in [0,1]$ be a fixed number. Then the following statements are all equivalent.*

(i) There is a transition T from $L(\mathscr{P}_1)$ to $L(\mathscr{P}_2)$ such that

$$\sup_{\theta \in \Theta} \|TP_\theta - Q_\theta\| \leq \varepsilon.$$

(ii) $\delta(\mathscr{P}_1, \mathscr{P}_2) \leq \varepsilon$.

(iii) For every finite set $S \subset \Theta$ one has $\delta(\mathscr{P}_{1,S}, \mathscr{P}_{2,S}) \leq \varepsilon$.

(iv) Let $\chi(\mu, \mathscr{P}_1, W)$ and $\chi(\mu, \mathscr{P}_2, W)$ be the characteristic envelope as defined in 2.3.4. Then, for each probability measure μ with finite support on Θ, each space $(A, \Gamma, W) \in \mathscr{C}$ one has

$$\chi(\mu, \mathscr{P}_1, W) \leq \chi(\mu, \mathscr{P}_2, W) + \varepsilon.$$

(v) Let H be a sublattice of the dual $M(\mathscr{P}_2)$ of $L(\mathscr{P}_2)$. Assume that $I \in H$ and that the unit ball of H is $\sigma(M(\mathscr{P}_2), L(\mathscr{P}_1))$-dense in the unit ball of $M(\mathscr{P}_2)$. For each probability measure μ with finite support on Θ, each space $(A, \Gamma, W) \in \mathscr{C}$, each (Γ, H)-continuous transition σ from $L(\mathscr{P}_2)$ to Γ', each $\varepsilon' > \varepsilon$, there is a transition ρ from $L(\mathscr{P}_1)$ to Γ' such that

$$R(\mu, \mathscr{P}_1, \rho, W, \Theta) \leq R(\mu, \mathscr{P}_2, \sigma, W, \Theta) + \varepsilon'.$$

Thus $\Delta(\mathscr{P}_1, \mathscr{P}_2) < \varepsilon$ means that for every procedure π_i in the problem i there is a procedure π_j in the problem j, $\{i, j\} = \{1, 2\}$, with risks differing by at most ε, uniformly over all bounded loss functions W and $\theta \in \Theta$.

For an experiment $\mathscr{P}_i$ and a regular decision space (A, W, Γ) there is a certain space $\mathscr{B}_i(\Gamma, W)$ of available decision procedures. This is the set of all transitions from $L(\mathscr{P}_i)$ to Γ^*, or equivalently the set of all positive normalized bilinear forms on $\Gamma \times L(\mathscr{P}_i)$.

Corollary 2.3.7 (Le Cam (1986))**.**

$$\delta(\mathscr{P}_1, \mathscr{P}_2) = \sup_{(A,W,\Gamma)} \sup_{\theta \in \Theta} \inf_{\pi_1 \in \mathscr{B}_1(\Gamma, W)} \sup_{\pi_2 \in \mathscr{B}_2(\Gamma, W)} \left| R(\mathscr{P}_1, \pi_1, W, \theta) - R(\mathscr{P}_2, \pi_2, W, \theta) \right|,$$

where the first supremum is taken over all regular decision spaces (A, W, Γ).

The introduction of the Δ-distance also makes it possible to look at the convergence of a sequence $(\mathscr{P}_{1,n})$ of experiments to a limit $\mathscr{P}_2$; more generally, it considers pairs $(\mathscr{P}_{1,n}, \mathscr{P}_{2,n})$ and ask whether $\Delta(\mathscr{P}_{1,n}, \mathscr{P}_{2,n}) \to 0$, as $n \to \infty$. Here the parameter space Θ needs not to be fixed: It can be Θ_n, depending on n. More precisely, for what concerns the terminology we have:

Definition 2.3.8. Two sequences of statistical models $(\mathscr{P}_{1,n})_n$ and $(\mathscr{P}_{2,n})_n$ are called *asymptotically equivalent* if $\Delta(\mathscr{P}_{1,n}, \mathscr{P}_{2,n})$ tends to zero as n goes to infinity.

2.4 Transitions and Markov kernels

From now on we specialize to the following more concrete setting. We consider a measurable space $(\mathscr{X}, \mathscr{T})$ called the *sample space*. Let $(P_\theta : \theta \in \Theta)$ be a family of probability measures on $(\mathscr{X}, \mathscr{T})$. We consider the L-space, endowed with the total variation norm, given by:

$$L = \Big\{\mu : \mu \text{ is a finite measure on } (\mathscr{X}, \mathscr{T}) \text{ s.t. } \mu \text{ is absolutely continuous} \\ \text{with respect to} \sum_{n=1}^{\infty} 2^{-n} P_{\theta_n} \text{ for a sequence } (\theta_n) \subseteq \Theta\Big\}$$

See Torgersen (1972) for more details.

Definition 2.4.1. Let μ and ν be two probability measures defined on a measurable space Ω. The *total variation distance* between μ and ν is defined as the quantity:

$$\|\mu - \nu\|_{TV} := \sup_{A \subset \Omega} |\mu(A) - \nu(A)|.$$

For a countable space Ω, the definition above becomes

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)|.$$

In this case it is easy to see that Markov kernels are in particular transitions.

However, the more general definition of transition has been given to insure that the decision procedures form a compact set on which the risk functions are lower semicontinuous. If, instead of the general definition of randomization given in Definition 2.2.10, we dealt only with Markov kernels, we would have lost this property. In other words, what we gain with this general setting is that the space of equivalence classes of experiments is a complete metric space.

However, there are many cases in which the set of all possible transitions $T : L(\mathscr{P}_1) \to L(\mathscr{P}_2)$ coincides with the set of Markov kernels $K : (\mathscr{X}_1, \mathscr{T}_1) \to (\mathscr{X}_2, \mathscr{T}_2)$. We list below some of these cases.

A first characterization in this sense can be found in Strasser (1985), Section 55, but this is not very relevant to our purposes since it assumes $\mathscr{X}_2$ locally compact. This would exclude, for example, the basic case $\mathscr{X}_2 = C_T$, the space of continuous functions on $[0, T]$. A more useful characterization concerns dominated statistical models with Polish sample spaces. We recall the following definitions:

Definition 2.4.2. A statistical model $\mathscr{P} = (\mathscr{X}, \mathscr{T}, (P_\theta)_{\theta\in\Theta})$ is called *Polish* if its sample space $(\mathscr{X}, \mathscr{T})$ is a separable completely metrizable topological space.

Example 2.4.3. Typical examples of Polish spaces in probability theory are the spaces $\mathbb{R}, \mathbb{R}^n, \mathbb{R}^\infty$, the space C_T of continuous functions on $[0, T]$ equipped with the supremum norm $d(x, y) = \sup_{0\leq t\leq T} |x_t - y_t|$, the space D of càdlàg functions equipped with the Skorohod metric (see Billingsley (1999)).

Definition 2.4.4. A statistical model $\mathscr{P} = (\mathscr{X}, \mathscr{T}, (P_\theta)_{\theta\in\Theta})$ is *dominated* if there exists a σ-finite measure μ on $(\mathscr{X}, \mathscr{T})$ such that, for all $\theta \in \Theta$, P_θ is absolutely continuous with respect to μ. The measure μ is said *dominating measure.*

Proposition 2.4.5. *If $\mathscr{P}_1$ is a dominated statistical model and $\mathscr{P}_2$ is a Polish one, then every transition from $L(\mathscr{P}_1)$ to $L(\mathscr{P}_2)$ is given by a Markov kernel.*

See Nussbaum (1996), page 2426 for the proof.

2.5 How to control the Le Cam distance

Even if the definition of deficiency has a perfectly reasonable statistical meaning, it is not easy to compute: Explicit computations have appeared but they are rare (see Hansen, Torgersen (1974); Torgersen (1972, 1974) and Section 1.9 in Shiryaev, Spokoiny (2000a)). One of these rare examples is the following (Torgersen (1974)). Let $\mathscr{P}_1$ be the experiment that observes $n + r$ i.i.d. variables, each distributed as a Gaussian random variable $\mathscr{N}(\theta, I)$, where I is the $k \times k$ identity matrix. Let $\mathscr{P}_2$ be the experiments that observes only the values of the first n variables. Then, according to Torgersen,

$$\Delta(\mathscr{P}_1, \mathscr{P}_2) = \mathbb{P}\bigg(\ln\frac{1+\alpha}{\alpha} \leq \frac{\chi^2}{k}\frac{1+\alpha}{\alpha}\ln(1+\alpha)\bigg),$$

where $\alpha = r/n$ and χ^2 has the central χ^2 distribution with k degrees of freedom.

More generally, one may hope to find more easily some upper bounds for the Δ-distance. We collect below some useful techniques for this purpose.

Property 2.5.1. *Let $\mathscr{P}_j = (\mathscr{X}, \mathscr{T}, \{P_{j,\theta}; \theta \in \Theta\})$, $j = 1, 2$, be two statistical models having the same sample space and define $\Delta_0(\mathscr{P}_1, \mathscr{P}_2) := \sup_{\theta\in\Theta} \|P_{1,\theta} - P_{2,\theta}\|_{TV}$. Then, $\Delta(\mathscr{P}_1, \mathscr{P}_2) \le \Delta_0(\mathscr{P}_1, \mathscr{P}_2)$.*

In particular, Property 2.5.1 allows us to bound the Δ-distance between statistical models sharing the same sample space by means of classical bounds for the total variation distance. To that aim, we collect below some useful (and classical) results.

In what follows, let μ and ν be two probability measures defined on a measurable space Ω. Let f and g denote their corresponding density functions with respect to a σ-finite dominating measure λ (for example, λ can be taken to be $\frac{\mu+\nu}{2}$). If $\Omega = \mathbb{R}$, let F, G denote their corresponding distribution functions. When needed, X, Y will denote random variables on Ω such that $\mathscr{L}(X) = \mu$ and $\mathscr{L}(Y) = \nu$. If Ω is a metric space, it will be understood to be a measurable space endowed with its Borel σ-algebra.

2.5.1 Total variation distance

Property 2.5.2. *Denoting by $L_1(\mu, \nu)$ the L_1 norm between μ and ν, i.e. $\int_\Omega |f(x) - g(x)|\lambda(dx)$, we have:*

$$\|\mu - \nu\|_{TV} = \frac{1}{2} L_1(\mu, \nu) = 1 - \int_\Omega \min(f(x), g(x))\mu(dx).$$

Proof. See, e.g., Tsybakov (2009), Lemma 2.1, page 84. □

Also recall that

$$L_1(\mu, \nu) = \max_{|h|\le 1} \Big| \int h d\mu - \int h d\nu \Big|,$$

where $h : \Omega \to \mathbb{R}$ satisfies $|h(x)| \le 1$, $\forall x \in \Omega$.

The total variation distance also has a coupling characterization (see Gibbs, Su (2002), p.19):

Property 2.5.3.

$$\|\mu - \nu\|_{TV} = \inf \big\{ \mathbb{P}(X \neq Y) : \text{ r.v. } X, Y \text{ s.t. } \mathscr{L}(X) = \mu,\ \mathscr{L}(Y) = \nu \big\}.$$

2.5.2 Hellinger distance and Hellinger process

Definition 2.5.4. The square of the *Hellinger distance* between μ and ν is defined as the quantity:

$$H^2(\mu,\nu) := \int_\Omega \Big(f(x)^{1/2} - g(x)^{1/2}\Big)^2 \lambda(dx).$$

This definition is independent of the choice of the dominating measure λ. For a countable space Ω,

$$H^2(\mu,\nu) := \sum_{\omega\in\Omega} \Big(\sqrt{\mu(\omega)} - \sqrt{\nu(\omega)}\Big)^2.$$

(See, e.g. Diaconis, Zabell (1982)). An important property is the following:

Property 2.5.5. *If μ and ν are product measures defined on the same measurable space, $\mu = \bigotimes_{j=1}^m \mu_j$ and $\nu = \bigotimes_{j=1}^m \nu_j$, then*

$$H^2(\mu,\nu) = 2\Bigg[1 - \prod_{j=1}^m \Bigg[1 - \frac{H^2(\mu_j,\nu_j)}{2}\Bigg]\Bigg].$$

(See, e.g., Zolotarev (1983), p. 279). Thus one can express the distance between distributions of vectors with independent components in terms of the component-wise distances. A consequence of Property 2.5.5 is:

Property 2.5.6. *If μ and ν are product measures defined on the same measurable space, $\mu = \bigotimes_{j=1}^m \mu_j$ and $\nu = \bigotimes_{j=1}^m \nu_j$, then*

$$H^2(\mu,\nu) \le 2\sum_{i=1}^m H^2(\mu_i,\nu_i).$$

(See, e.g., Strasser (1985), Lemma 2.19).

Property 2.5.7. *The Hellinger distance between two normal distributions $\mu \sim \mathcal{N}(m_1,\sigma_1^2)$ and $\nu \sim \mathcal{N}(m_2,\sigma_2^2)$ is:*

$$H^2(\mu,\nu) = 2\Bigg[1 - \Bigg[\frac{2\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2}\Bigg]^{1/2} \exp\Bigg[-\frac{(m_1-m_2)^2}{4(\sigma_1^2+\sigma_2^2)}\Bigg]\Bigg].$$

Finally, two properties of the Hellinger distance often used to bound Δ_0 are:

Property 2.5.8. *For all measures μ and ν*

$$L_1(\mu,\nu) \le 2H(\mu,\nu).$$

Property 2.5.9. *For all measures μ and ν*

$$\frac{H^2(\nu,\mu)}{2} \leq \|\nu-\mu\|_{TV} \leq H(\nu,\mu)\sqrt{1-\frac{H^2(\nu,\mu)}{4}}.$$

Proof. See, e.g., Tsybakov (2009), Lemma 2.3, page 86. □

Let us now focus on the Hellinger process, that allows us to derive a useful formula for looking for equivalence results for continuous processes.

Let $(\Omega,\mathscr{A})$ be a measurable space with a right-continuous filtration $A=(\mathscr{A}_t)_{t\geq 0}$, $\mathscr{A}=\bigvee_{t\geq 0}\mathscr{A}_t$, and probability measures P, $\tilde{P}$ and $Q:=\frac{P+\tilde{P}}{2}$. For convenience of formulations we assume that the space $(\Omega,\mathscr{A},Q)$ is complete and $\mathscr{A}_0$ contains all Q-null sets from $\mathscr{A}$. $\mathcal{T}$ is the set of all stopping times with respect to A. We denote by P_T the restriction of P to the σ-algebra $\mathscr{A}_T$.

Let $z=(z_t)_{t\geq 0}$ be the density process of the measure P with respect to Q, that is to say $z_t=\frac{dP_t}{dQ_t}$, for all $t\geq 0$. Note that z, for any $T\in\mathcal{T}$ and $\omega\in D$, satisfies:

$$z_T(\omega)=\frac{dP_T}{dQ_T}(\omega).$$

Similarly, consider the density process $\tilde{z}$ of the probability measure $\tilde{P}$ with respect to Q.

Definition 2.5.10. Let $0<\alpha<1$. Define $Y_t^{(\alpha)}=z_t^\alpha\tilde{z}_t^{1-\alpha}$. The *$\alpha$-Hellinger process* is the only càdlàg predictable increasing process $h^{(\alpha)}$ such that $h_0^{(\alpha)}=0$ and

$$Y_t^{(\alpha)}+\int_0^t Y_{s-}^{(\alpha)}dh_s^{(\alpha)}$$

is a Q-martingale.

Let R be an arbitrary probability measure dominating P as well as $\tilde{P}$. The *Hellinger integral of order α* is defined by

$$H(\alpha,P,\tilde{P})=\mathbb{E}_R\bigg(\Big(\frac{dP}{dR}\Big)^\alpha\Big(\frac{d\tilde{P}}{dR}\Big)^{1-\alpha}\bigg).$$

The value $H(\alpha,P,\tilde{P})$ does not depend on the choice of R.

Lemma 2.5.11. *For any $\alpha\in(0,1)$*

$$2(1-H(\alpha,P,\tilde{P}))\leq\|P-\tilde{P}\|_{TV}\leq\sqrt{C_\alpha(1-H(\alpha,P,\tilde{P}))},$$

where C_α is a constant; for $\alpha=\frac{1}{2}$ it is possible to choose $C_{\frac{1}{2}}=8$.

Besides,

$$H(\alpha, P, \tilde{P}) = \mathbb{E}_Q\Big(z_T^{\alpha}\tilde{z}_T^{1-\alpha}\Big), \quad \forall T \in \mathcal{T}.$$

Put $d_0 = \|P_0 - \tilde{P}_0\|_{TV}$. Two useful bounds for the total variation distance between P_T and $\tilde{P}_T$ based only on the value of the Hellinger process $h^{(1/2)}$ at time $T \in \mathcal{T}$ are the following:

Theorem 2.5.12. *Let $T \in \mathcal{T}$, $\varepsilon > 0$. Then:*

$$\|P_T - \tilde{P}_T\|_{TV} \le d_0 + 4\sqrt{\mathbb{E}_P h_T^{(1/2)}},$$
$$\|P_T - \tilde{P}_T\|_{TV} \le \frac{3}{2}d_0 + 3\sqrt{2\varepsilon} + 2P\Big(h_T^{(1/2)} \ge \varepsilon\Big).$$

A proof of Lemma 2.5.11 and Theorem 2.5.12 can be found in Jacod, Shiryaev (1987).

2.5.3 κ- distance

Definition 2.5.13. Denote by h the density $h(x) = \frac{d(\mu+\nu)}{d\lambda}(x)$. The κ-*distance* between μ and ν is defined as:

$$\kappa^2(\mu, \nu) = \frac{1}{2}\int \frac{(f(x) - g(x))^2}{h(x)}\lambda(dx).$$

Property 2.5.14. *We have:*

$$\frac{1}{2}H^2(\mu, \nu) \le \kappa^2(\mu, \nu) \le H^2(\mu, \nu),$$

hence, by means of Property 2.5.2,

$$\kappa^2(\mu, \nu) \le 2\|\mu - \nu\|_{TV}.$$

Proof. In order to prove that $\frac{1}{2}H^2(\mu, \nu) \le \kappa^2(\mu, \nu)$ write:

$$\kappa^2(\nu, \mu) = \frac{1}{2}\int_\Omega \big(\sqrt{f(x)} - \sqrt{g(x)}\big)^2 \frac{\big(\sqrt{f(x)} + \sqrt{g(x)}\big)^2}{h(x)}\lambda(dx)$$

and observe that

$$h(x) \le \big(\sqrt{f(x)} + \sqrt{g(x)}\big)^2 \le 2h(x).$$

The inequality $\kappa^2(\mu, \nu) \le H^2(\mu, \nu)$ follows by Property 2.5.2. $\square$

2.5.4 Relative entropy (or Kullback-Leibler divergence)

Definition 2.5.15. The *relative entropy* between μ and ν is defined as the quantity:

$$K(\mu,\nu) = \begin{cases} \int_\Omega f(x) \log\left(\frac{f}{g}(x)\right)\lambda(dx) & \text{if} \quad \mu \ll \nu, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\mu \ll \nu$ stands for "μ is absolutely continuous with respect to ν".

The definition is independent of the choice of the dominating measure λ. For Ω a countable space,

$$K(\mu,\nu) := \sum_{\omega\in\Omega} \mu(\omega)\log\left(\frac{\mu(\omega)}{\nu(\omega)}\right).$$

Relative entropy is not a metric, since it is not symmetric and does not satisfy the triangle inequality. However, it has many useful properties, including additivity over marginals of product measures (see Reiss (1989)):

Property 2.5.16. *If μ and ν are product measures defined on the same measurable space, $\mu = \bigotimes_{j=1}^m \mu_j$ and $\nu = \bigotimes_{j=1}^m \nu_j$, then*

$$K(\mu,\nu) = \sum_{i=1}^m K(\mu_i,\nu_i).$$

A useful property for our purposes is the following:

Proposition 2.5.17.

$$H^2(\mu,\nu) \le K(\mu,\nu).$$

Remark 2.5.18. Relative entropy was first defined by Kullback, Leibler (1951) as a generalization of the entropy notion of Shannon (1948). A standard reference for its properties is Cover, Thomas (2012).

2.5.5 χ^2-distance

Definition 2.5.19. Denote by $S(\mu)$, $S(\nu)$ the supports on Ω of the measures μ, ν. The χ^2-*distance* between μ and ν is defined as the quantity:

$$\chi^2(\mu,\nu) = \int_{S(\mu)\cup S(\nu)} \frac{(g(x)-f(x))^2}{g(x)}\lambda(dx).$$

The definition is independent of the choice of the dominating measure λ. For Ω a countable space, the definition of the χ^2-distance reduces to:

$$\chi^2(\mu,\nu)=\sum_{\omega\in S(\mu)\cup S(\nu)}\frac{\big(\mu(\omega)-\nu(\omega)\big)^2}{\nu(\omega)}.$$

The χ^2-distance is not symmetric and therefore it is not a metric. However, like the Hellinger distance and the relative entropy, the χ^2-distance between product measures can be bounded in terms of the distances between their marginals. See Reiss (1989), p. 100. A useful property is the following

Proposition 2.5.20.

$$\chi^2(\mu,\nu)\leq 2\kappa^2(\mu,\nu).$$

2.5.6 The likelihood process

Another way to control the Le Cam distance lies in the deep relation linking the equivalence between experiments to the proximity of the distributions of the related likelihood ratios.

Let $\mathscr{P}_j=(\mathscr{X}_j,\mathscr{T}_j,(P_{j,\theta})_{\theta\in\Theta})$ be a statistical model dominated by P_{j,θ_0}, $\theta_0\in\Theta$, and let $\Lambda_j(\theta)=\frac{dP_{j,\theta}}{dP_{j,\theta_0}}$ be the density of $P_{j,\theta}$ with respect to P_{j,θ_0}. In particular, one can see $\Lambda_j(\theta)$ as a real random variable defined on the probability space $(\mathscr{X}_j,\mathscr{T}_j)$, i.e. one can see $(\Lambda_j(\theta))_{\theta\in\Theta}$ as a stochastic process. For that reason we introduce the notation $\Lambda_{\mathscr{P}_j}:=(\Lambda_j(\theta),\theta\in\Theta)$ and we call $\Lambda_{\mathscr{P}_j}$ the *likelihood process.*

A key result of the theory of Le Cam is the following.

Property 2.5.21. *Let $\mathscr{P}_j=(\mathscr{X}_j,\mathscr{T}_j,(P_{j,\theta})_{\theta\in\Theta})$, $j=1,2$ be two experiments. If the family $(P_{j,\theta})$ is dominated by $(P_{j,\theta_0})_{\theta\in\Theta}$, then $\mathscr{P}_1$ and $\mathscr{P}_2$ are equivalent if and only if their likelihood processes under the dominating measures P_{1,θ_0} and P_{2,θ_0} coincide.*

Proof. see Strasser (1985), Corollary 25.9. □

Let us now suppose that there are two processes $(\Lambda_j^{n,*}(\theta))_{\theta\in\Theta}$, $j=1,2$ defined on a same probability space $(\mathscr{X}^*,\mathscr{T}^*,\Pi^*)$ and such that the law of $(\Lambda_j^n(\theta))_{\theta\in\Theta}$ under P_{j,θ_0} is equal to the law of $(\Lambda_j^{n,*}(\theta))_{\theta\in\Theta}$ under Π^*, $j=1,2$. Then, the following holds (see Le Cam, Yang (2000), Lemma 6).

Property 2.5.22. *If $\Lambda_{\mathscr{P}_1}$ and $\Lambda_{\mathscr{P}_2}$ are the likelihood processes associated with the experiments $\mathscr{P}_1$ and $\mathscr{P}_2$, then*

$$\Delta(\mathscr{P}_1^n,\mathscr{P}_2^n)\leq\sup_{\theta\in\Theta}\mathbb{E}_{\Pi^*}\Big|\Lambda_1^{n,*}(\theta)-\Lambda_2^{n,*}(\theta)\Big|.$$

2.5.7 Sufficiency and Le Cam distance

A very useful tool, when comparing statistical models having different sample spaces, is to look for a sufficient statistics. The introduction of the term *sufficient statistics* is usually attributed to R.A. Fisher who gave several definitions of the concept. We cite here Le Cam's presentation of the subject from Le Cam (1964). Fisher's most relevant statement seems to be the requirement "...that the statistic chosen should summarize the whole of the relevant information supplied by the sample." Such a requirement may be made precise in various ways, the following three interpretations are the most common.

(i) *The classical, or operational definition of sufficiency* claims that a statistics S is sufficient if, given the value of S, one can proceed to a post-experimental randomization reproducing variables which have the same distributions as the originally observable variables.

(ii) *The Bayesian interpretation.* A statistics S is sufficient if for every a priori distribution of the parameter the a posteriori distributions of the parameter given S is the same as if the entire result of the experiment was given.

(iii) *The economic interpretation.* A statistics S is sufficient if for every decision problem and every decision procedure made available by the experiment there is a decision procedure, depending on S only, which has the same performance characteristics.

Formally, let $\mathscr{P}_1 = (\mathscr{X}_1, \mathscr{T}_1, \{P_{1,\theta}, \theta \in \Theta\})$ be a statistical model dominated by the measure μ. Let $S : (\mathscr{X}_1, \mathscr{T}_1) \to (\mathscr{X}_2, \mathscr{T}_2)$ be a random variable.

Definition 2.5.23. S is said a *sufficient statistics* is for any bounded random variable Y on $\mathscr{X}_1$ there exists a version of the conditional expectation $\mathbb{E}^\theta[Y|S]$ not depending on θ and defined μ-a.e.

An important result linking the Le Cam distance with the existence of a sufficient statistic is the following (see e.g. Fact 1.7 in Mariucci (2015b)):

Property 2.5.24. *Let $\mathscr{P}_i = (\mathscr{X}_i, \mathscr{A}_i, \{P_{i,\theta}, \theta \in \Theta\})$, $i = 1, 2$, be two statistical models. Let $S : \mathscr{X}_1 \to \mathscr{X}_2$ be a sufficient statistics such that the distribution of S under $P_{1,\theta}$ is equal to $P_{2,\theta}$. Then $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$.*

2.6 The current state of research

Several results appeared in the literature concerning nonparametric asymptotic equivalences for statistical models. In the following, we decided to describe the main results of a selection of those. This detailed description is followed by a more inclusive bibliographical list of results on the subject.

Brown, Low (1996): They consider the problem of estimating the function f from a continuously observed Gaussian process $y(t)$, $t \in [0,1]$, which satisfies the SDE

$$dy_t = f(t)dt + \frac{\sigma(t)}{\sqrt{n}}dW_t, \quad t \in [0,1],$$

where dW_t is a Gaussian white noise. They find that the statistical model associated with the continuous observation of (y_t) is asymptotically equivalent to the statistical model associated with its discrete counterpart, i.e. the nonparametric regression:

$$y_i = f(t_i) + \sigma(t_i)\xi_i, \quad i = 1, \dots, n.$$

The time grid is uniform, $t_i = \frac{i-1}{n}$, and the ξ_i's are standard normal variables; they assume that f varies in a nonparametric subset $\mathscr{F}$ of $L_2[0,1]$ defined by some smoothness conditions and n tends to infinity not too slowly. More precisely, the drift function $f(\cdot)$ is unknown and such that, for B a positive constant, one has:

$$\sup\big\{|f(t)| : t \in [0,1], f \in \mathscr{F}\big\} = B < \infty.$$

Moreover, defining

$$\bar{f}_n(t) = \begin{cases} f\left(\frac{i}{n}\right) & \text{if} \quad \frac{i-1}{n} \leq t < \frac{i}{n}, \quad i = 1, \dots, n; \\ f(1) & \text{if} \quad t = 1; \end{cases}$$

one asks:

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} n \int_0^1 \frac{(f(t) - \bar{f}_n(t))^2}{\sigma^2(t)} dt = 0.$$

The diffusion coefficient $\sigma^2(\cdot) > 0$ is supposed to be a known absolutely continuous function on $[0,1]$ such that

$$\left|\frac{d}{dt}\ln\sigma(t)\right| \leq C, \quad t \in [0,1],$$

for some positive constant C.

Nussbaum (1996): In this paper Nussbaum establishes a global asymptotic equivalence between the problem of density estimation from an i.i.d. sample and a Gaussian white noise model. More precisely, let $(Y_i)_{i=1}^n$ be i.i.d. random variables with density f on $[0,1]$ with respect to the Lebesgue measure. The densities f are the unknown parameters and they are supposed to belong to a certain nonparametric class $\mathscr{F}$ subject to a Hölder restriction: $|f(x) - f(y)| \leq C|x - y|^\alpha$ with $\alpha > \frac{1}{2}$ and a positivity restriction: $f(x) \geq \varepsilon > 0$. Let us denote by $\mathscr{P}_{1,n}$ the statistical model associated with the observation of the Y_i's. Furthermore, let $\mathscr{P}_{2,n}$ be the experiment in which one observes a stochastic process $(Y_t)_{t\in[0,1]}$ such that

$$dY_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{n}}dW_t, \quad t \in [0, 1],$$

where $(W_t)_{t\in[0,1]}$ is a standard Brownian motion. Then the main result in Nussbaum (1996) is that $\Delta(\mathscr{P}_{1,n}, \mathscr{P}_{2,n}) \to 0$ as $n \to \infty$.

Milstein, Nussbaum (1998): The authors consider the problem of estimating the function f from a continuously or discretely observed diffusion process $y(t)$, $t \in [0, 1]$, which satisfies the SDE:

$$dy_t = f(y_t)dt + \varepsilon dW_t, \quad t \in [0, 1], \quad y_0 = 0,$$

where (W_t) is a standard Brownian motion and ε is a known small parameter.

The drift function $f(\cdot)$ is unknown and such that, for K a positive constant,

$$f \in \mathscr{F}_K = \Big\{f \text{ defined on } \mathbb{R} \text{ and } \forall x, u \in \mathbb{R},\ |f(x) - f(u)| \leq K|x - u|,\ |f(0)| \leq K\Big\}.$$

The constant K has to exist but may be unknown. For what concerns the discrete observation of (y_t) the authors place themselves in a high-frequency framework: $t_i = \frac{i}{n}$, $i \leq n$. Their main result is that, if $n\varepsilon \to \infty$ as $\varepsilon \to 0$, then there is an asymptotic equivalence between the continuous observation of (y_t) and the corresponding Euler scheme:

$$Z_0 = 0, \quad Z_i = Z_{i-1} + \frac{f(Z_{i-1})}{n} + \frac{\varepsilon}{\sqrt{n}}\xi_i, \quad i = 1, \dots, n,$$

where (ξ_i) are i.i.d. standard normal variables. Denoting by $\mathscr{P}$ and $\mathscr{Z}_n$ the statistical models associated with the continuous observation of (y_t) and the Euler scheme, respectively, an upper bound for the rate of convergence of $\Delta(\mathscr{P}, \mathscr{Z}_n)$ is given by:

$$\Delta(\mathscr{P}, \mathscr{Z}_n) = O\Big(\sqrt{n^{-2}\varepsilon^{-2} + n^{-1}}\Big), \quad \text{as } \varepsilon \to 0.$$

The authors also prove that the discrete observations $(y_{t_1}, \dots, y_{t_n})$ form an asymptotically sufficient statistics.

Carter (2002): In this paper Carter establishes a global asymptotic equivalence between a density estimation model and a Gaussian white noise model by bounding the Le Cam distance between multinomial and multivariate normal random variables. More precisely, let us denote by $\mathcal{M}(n,\theta)$ the multinomial distribution, where $\theta := (\theta_1,\dots,\theta_m)$. Denote the covariance matrix nV_θ: Its (i,j)th element equals to $n\theta_i(1-\theta_i)\delta_{i,j} - n\theta_i\theta_j$.

The main result is an upper bound for the Le Cam distance $\Delta(\mathcal{M},\mathcal{N})$ between the models $\mathcal{M} := \{\mathcal{M}(n,\theta) : \theta \in \Theta\}$ and $\mathcal{N} := \{\mathscr{N}(n\theta, nV_\theta) : \theta \in \Theta\}$, under some regularity assumptions on Θ. In particular, Carter proves that

$$\Delta(\mathcal{M},\mathcal{N}) \leq C'_\Theta \frac{m \ln m}{\sqrt{n}} \quad \text{provided} \quad \sup_{\theta\in\Theta} \frac{\max_i \theta_i}{\min_i \theta_i} \leq C_\Theta < \infty,$$

for a constant C'_Θ that depends only on C_Θ. From this inequality Carter can recover the same results as Nussbaum (1996) under stronger regularity assumptions on $\mathscr{F}$: Here $\mathscr{F}$ is a class of smooth, differentiable densities f on the interval $[0,1]$ such that there exist strictly positive constants ε, M, γ such that $\varepsilon \leq f \leq M$ and

$$|f'(x) - f'(y)| \leq M|x-y|^\gamma, \quad \text{for all } x,y \in [0,1].$$

Carter (2007): In this paper Carter shows that a nonparametric regression experiment with drift problem where the variance is unknown is asymptotically equivalent to a continuous Gaussian process. In particular, he proves that there is no significant penalty to pay in treating the variance as part of the parameter space because the equivalence holds for essentially the same space as in Brown, Low (1996). The considered regression model is associated with the observations of the Y_i's random variables:

$$Y_i = f\Big(\frac{i}{n}\Big) + \sigma\xi_i, \quad i = 1,\dots,n,$$

with standard normal variables ξ_i. The parameter space is of the form $\mathscr{F}_\sigma \times \mathbb{R}^+$ in order to include both the smooth function f and the variance σ^2. We will not give the exact definition of the parameter set for the sake of brevity and because it would lead us to a detour in the subject of wavelet decompositions. We only mention that the parameter space $\mathscr{F}_\sigma$ includes all functions in a Hölder space with $\alpha > \frac{1}{2}$ as in Brown, Low (1996). The main result is the following: Observing $(Y_i)_{i=1}^n$ is asymptotically equivalent to the statistical model that observes the continuous process

$$Y_t|V = \int_0^t f(x)dx + \frac{\sqrt{V}}{\sqrt{n}} W_t, \quad 0 \leq t \leq 1,$$

where (W_t) is a standard Brownian motion and V is a Gamma random variable $\Gamma\big(\frac{n}{2}, \frac{2\sigma^2}{n}\big)$.

Genon-Catalot, Laredo (2014): The authors consider the diffusion process (ξ_t) given by

$$d\xi_t = b(\xi_t)dt + \sigma(\xi_t)dW_t, \quad \xi_0 = \eta, \tag{2.1}$$

where (W_t) is a standard Brownian motion defined on a filtered probability space $(\Omega, \mathscr{A}, (\mathscr{A}_t)_{t\geq 0}, \mathbb{P})$, η is a real valued $\mathscr{A}_0$-measurable random variable, $b(\cdot)$, $\sigma(\cdot)$ are real-valued functions defined on $\mathbb{R}$. The diffusion coefficient $\sigma(\cdot)$ is a known nonconstant function that belongs to $C^2(\mathbb{R})$ and satisfies the conditions:

$$\forall x \in \mathbb{R},\ 0 < \sigma_0^2 \leq \sigma^2(x) \leq \sigma_1^2, \quad |\sigma'(x)| + |\sigma''(x)| \leq K_\sigma.$$

The drift function $b(\cdot)$ is unknown and such that, for K a positive constant,

$$b(\cdot) \in \mathscr{F}_K = \Big\{ b(\cdot) \in C^1(\mathbb{R}) \text{ and for all } x \in \mathbb{R},\ |b(x)| + |b'(x)| \leq K \Big\}.$$

The constant K has to exist but may be unknown. The sample path of (ξ_t) is continuously and discretely observed on a time interval $[0,T]$. The discrete observations of (ξ_t) occur at the times $t_i = ih$, $i \leq n$ with $T = nh$. The authors prove the asymptotic equivalence between the continuous or discrete observation of (ξ_t) and the corresponding Euler scheme:

$$Z_0 = \eta, \quad Z_i = Z_{i-1} + hb(Z_{i-1}) + \sqrt{h}\sigma(Z_{i-1})\varepsilon_i,$$

where, for $i \geq 1$, $t_i = ih$ and $\varepsilon_i = \frac{W_{t_i} - W_{t_{i-1}}}{\sqrt{h}}$. The equivalences hold under the assumptions that n tends to infinity with $h = h_n$ and $nh_n^2 = \frac{T^2}{n}$ tends to zero. This includes both the case $T = nh_n$ bounded and the one $T \to \infty$. Let us stress the rate of convergence in the small variance case, that is obtained by replacing σ by $\varepsilon\sigma$ in (2.1). Let us also denote by $\mathscr{P}$ and $\mathscr{Z}_n$ the statistical models associated with the continuous observation of (ξ_t) and the Euler scheme, respectively. The computations in Genon-Catalot, Laredo (2014) give the following upper bound for the rate of convergence:

$$\Delta(\mathscr{P}, \mathscr{Z}_n) = O\Big(\sqrt{n^{-2}\varepsilon^{-2} + n^{-1} + n^{-1}\varepsilon^{-4}}\Big).$$

We conclude this section with a broader collection of other equivalence results for nonparametric experiments. Here, we will limit ourselves to citing the relevant contributions, the reader is referred to the original papers for more details. Asymptotic equivalence theory has been developed for nonparametric regression in Brown, Low (1996); Brown et al. (2002b); Carter (2006b, 2007, 2009); Grama, Nussbaum (2002); Meister, Reiß (2013); Reiß (2008); Rohde (2004), nonparametric density estimation models in Brown et al. (2004a); Carter (2002); Jähnisch, Nussbaum (2003); Nussbaum (1996), generalized linear models in Grama, Nussbaum (1998), time series in Grama, Neumann (2006); Milstein, Nussbaum (1998), diffusion models in Dalalyan, Reiß (2006, 2007b); Delattre, Hoffmann (2002); Genon-Catalot, Laredo (2014); Genon-Catalot, Laredo, Nussbaum (2002); Mariucci (2015c); Reiß (2011), functional linear regression in Meister (2011), spectral density estimation in Golubev, Nussbaum, Zhou (2010) and processes with jumps in Mariucci (2015b). It is generally harder to prove negative results, that is, that two given models are not equivalent; the most notable results in this direction are Brown, Zhang (1998); Efromovich, Samarov (1996); Wang (2002a).

Chapter 3

Asymptotic equivalence for inhomogeneous jump diffusion processes and white noise

Résumé Dans le Chapitre 3 nous présentons un résultat d'équivalence asymptotique entre les expériences associées à l'observation discrète (haute fréquence) ou continue d'un processus additif (avec mesure de Lévy finie) et un modèle de bruit blanc gaussien. Ici, le paramètre d'intérêt est la fonction de dérive et le temps d'observation peut être fini ou infini. Ces équivalences asymptotiques sont établies en construisant des noyaux de Markov explicites qui peuvent être utilisés pour reproduire une expérience à partir de l'autre. Ce chapitre est basé sur un article publié dans *ESAIM: Probability and Statistics*, Mariucci 2015b.

Mot clés: Expériences statistiques non paramétriques, distance de Le Cam, processus de Lévy, processus additifs, modèle de bruit blanc gaussien.

Abstract We prove the global asymptotic equivalence between the experiments generated by the discrete (high frequency) or continuous observation of a path of a time inhomogeneous jump-diffusion process and a Gaussian white noise experiment. Here, the parameter of interest is the drift function and the observation time T can be both bounded or diverging. The approximation is given in the sense of the Le Cam distance, under some smoothness conditions on the unknown drift function. These asymptotic equivalences are established by constructing explicit Markov kernels that can be used to reproduce one experiment from the other. Chapter 3 is based on a paper published in *ESAIM: Probability*

and Statistics.

Keywords: Nonparametric experiments, Le Cam distance, Lévy processes, additive processes, white noise.

3.1 Introduction

Consider a sequence of one-dimensional time inhomogeneous jump-diffusion processes $\{X_t\}_{t\geq 0}$ defined by

$$X_t = \eta + \int_0^t f(s)ds + \int_0^t \sigma_n(s)dW_s + \sum_{i=1}^{N_t} Y_i, \quad t \in [0, T_n], \tag{3.1}$$

where:

- η is some random initial value;
- $W = \{W_t\}_{t\geq 0}$ is a standard Brownian motion;
- $N = \{N_t\}_{t\geq 0}$ is an inhomogeneous Poisson process independent of W and with intensity function $\lambda(\cdot)$;
- $(Y_i)_{i\geq 1}$ is a sequence of i.i.d. real random variables with distribution G, independent of W and N;
- $\sigma_n^2(\cdot)$ is supposed to be known. Either $T_n \to \infty$ and $\sigma_n(\cdot) = \sigma(\cdot)$ does not depend on n or $T_n \equiv T$ and $\sigma_n(\cdot) = \varepsilon_n \sigma(\cdot)$ with $\varepsilon_n \to 0$ as $n \to \infty$.
- $f(\cdot)$ belongs to some non-parametric class $\mathscr{F}$ making its estimation consistent (e.g. if $T_n \to \infty$ one may require $\mathscr{F}$ to consist of a subclass of periodic functions).
- $\lambda(\cdot)$ and $G(\cdot)$ are unknown and belong to non-parametric classes Λ and $\mathscr{G}$, respectively.

We observe $\{X_t\}_{t\geq 0}$ at discrete times $0 = t_0 < t_1 < \cdots < t_n = T_n$ such that $\Delta_n = \max_{1\leq i\leq n} \big\{|t_i - t_{i-1}|\big\} \downarrow 0$ as n goes to infinity. We are interested in estimating the drift function $f(\cdot)$ from the discrete data $(X_{t_i})_{i=0}^n$. At least two natural questions arise:

1. How much information about the parameter $f(\cdot)$ do we lose by observing $(X_{t_i})_{i=0}^n$ instead of $\{X_t\}_{t\in[0,T_n]}$?

2. Can we construct an easier (read: mathematically more tractable), but equivalent, model from $(X_{t_i})_{i=0}^n$?

The aim of this paper is to give an answer to questions (1) and (2) by means of the Le Cam theory of statistical experiments. For the basic concepts and a detailed description of the notion of asymptotic equivalence that we shall adopt, we refer to Le Cam (1986); Le Cam, Yang (2000). We recall the relevant definitions and properties in Section 3.2.2.

One of the main applications of proving an asymptotic equivalence between two sequences of experiments is that it allows to transfer asymptotic risk bounds for any inference problem from one model to the other, at least for bounded loss functions. In particular, if there is an estimator τ_1 in the statistical model $\mathscr{P}_1 = (\mathscr{X}_1, \mathscr{A}_1, \{P_{1,\theta} : \theta \in \Theta\})$ with risk $\int L(\theta, \tau(x)) P_{1,\theta}(dx)$, then, for bounded loss functions L, there is an estimator τ_2 in $\mathscr{P}_2$ such that

$$\sup_\theta \left| \int L(\theta, \tau_1(x)) P_{1,\theta}(dx) - \int L(\theta, \tau_2(x)) P_{2,\theta}(dx) \right| \to 0, \quad \text{as } n \to \infty.$$

More generally, asymptotic equivalence allows to transfer minimax rates of convergence, up to some constants.

The first asymptotic equivalence results for non-parametric experiments date to 1996 and are due to Brown, Low (1996) and Nussbaum (1996). This is the first instance of an abundance of works devoted to establishing asymptotic equivalence results for non-parametric experiments. In particular, asymptotic equivalence theory has been developed for non-parametric regression Brown, Low (1996); Brown et al. (2002b); Carter (2006b, 2007, 2009); Grama, Nussbaum (2002); Meister, Reiß (2013); Reiß (2008); Rohde (2004), non-parametric density estimation models Brown et al. (2004a); Carter (2002); Jähnisch, Nussbaum (2003); Nussbaum (1996), generalized linear models Grama, Nussbaum (1998), time series Grama, Neumann (2006); Milstein, Nussbaum (1998), diffusion models Dalalyan, Reiß (2006, 2007b); Delattre, Hoffmann (2002); Genon-Catalot, Laredo (2014); Genon-Catalot, Laredo, Nussbaum (2002); Mariucci (2015c); Reiß (2011), GARCH model Buchmann, Müller (2012), functional linear regression Meister (2011) and spectral density estimation Golubev, Nussbaum, Zhou (2010). Negative results are somewhat harder to come by; the most notable among them are Brown, Zhang (1998); Efromovich, Samarov (1996); Wang (2002a).

There is however a lack of equivalence results concerning processes with jumps. To our knowledge, this is the first one for what concerns the estimation of a drift function issued from a discretely (high frequency) observed Lévy process. We actually allow it to be inhomogeneous in time, i.e. an additive process. In this setting one should also cite the works Duval, Hoffmann (2011); Etoré, Louhichi, Mariucci (2013) as they are the only ones we know about treating (pure jumps) Lévy processes. However, they both give asymptotic results for the estimation of the Lévy measure.

The interest in Lévy processes is due to them being a building block for stochastic continuous time models with jumps. Because of that, they are widely used in finance, queueing, telecommunications, extreme value theory, quantum theory or biology. Their stationarity property, however, makes them rather inflexible; as a consequence, in recent years additive processes have been preferred in financial modelling (see Cont, Tankov (2004), Chapter 14). It is therefore in this more general setting that we present our results.

In order to mathematically reformulate questions (1) and (2), let us denote by $(D, \mathscr{D})$ the Skorokhod space; define $P_{T_n}^{(f,\sigma_n^2,\lambda G)}$ as the law of $\{X_t\}_{t\in[0,T_n]}$ on $(D,\mathscr{D})$ and $Q_n^{(f,\sigma_n^2,\lambda G)}$ as the law of the vector $(X_{t_0}, X_{t_1}, \dots, X_{t_n})$ on $(\mathbb{R}^{n+1}, \mathscr{B}(\mathbb{R}^{n+1}))$.

Consider the parameter set $\Theta = \mathscr{F}$. We allow two more degrees of freedom by considering $\lambda \in \Lambda$ and $G \in \mathscr{G}$, although these will not be parameters of interest. Let us then consider the following statistical models:

$$\mathscr{P}_n = \big(D, \mathscr{D}, \{P_{T_n}^{(f,\sigma_n^2,\lambda G)} : f \in \mathscr{F}\}\big),$$
$$\mathscr{Q}_n = \big(\mathbb{R}^{n+1}, \mathscr{B}(\mathbb{R}^{n+1}), \{Q_n^{(f,\sigma_n^2,\lambda G)} : f \in \mathscr{F}\}\big).$$

Finally, let us introduce the Gaussian model that will appear in the statement of our main results. For that, let us denote by $(C, \mathscr{C})$ the space of continuous mappings from $[0,\infty)$ into $\mathbb{R}$ endowed with its standard filtration and, coherently with the previous notation, by $P_{T_n}^{(f,\sigma_n^2,0)}$ the law induced on $(C,\mathscr{C})$ by the stochastic process:

$$dy_t = f(t)dt + \sigma_n(t)dW_t, \quad y_0 = 0, \quad t \in [0, T_n]. \tag{3.2}$$

We set:

$$\mathscr{W}_n = \big(C, \mathscr{C}, (P_{T_n}^{(f,\sigma_n^2,0)} : f \in \mathscr{F})\big).$$

We have already mentioned that asymptotic equivalences can be used to reduce estimation problems from one model to a simpler ones. This is what happens here, the model associated with the discrete or continuous observation of $\{X_t\}$ as in (3.1) has been proved to be equivalent to that in (3.2), which is much better studied. For example, consider the two following situations:

- T_n is fixed and $\sigma_n(\cdot) = \varepsilon_n \sigma(\cdot)$ with $\varepsilon_n \to 0$,
- T_n goes to infinity and $\sigma_n(\cdot)$ is fixed; in this case, also ask that elements of $\mathscr{F}$ have some periodicity assumption.

In both these cases, a consistent estimation of $f \in \mathscr{F}$ is possible. Our equivalence result does not rely on assumptions such as these, but it applies to these cases, as well: Indeed, proving equivalence for a class $\mathscr{F}$ automatically implies that the same equivalence holds true for any subclass of $\mathscr{F}$.

We state here our main result in the case in which $\mathscr{F}$ is a functional class consisting of α-Hölder, uniformly bounded functions on $\mathbb{R}$, i.e. there exist $B < \infty$, $M < \infty$ and $\alpha \in (0,1]$ such that

$$|f(x)| \leq B \text{ and } |f(x) - f(y)| \leq M|x-y|^\alpha, \quad \forall x, y \in \mathbb{R}.$$

For the general statements see Section 3.2.4.

Theorem 3.1.1. *Suppose that $\mathscr{F}$ is a subclass of α-Hölder, uniformly bounded functions on $\mathbb{R}$. Let $\sigma_n(\cdot) = \varepsilon_n \sigma(\cdot)$ be such that $0 < m_\sigma \leq \sigma(\cdot) \leq M_\sigma < \infty$ with derivative $\sigma'(\cdot)$ in $L_\infty(\mathbb{R})$. Suppose either:*

- *$T_n \equiv T < \infty$, $\varepsilon_n \to 0$ and there exists an $L_2 < \infty$ such that for all $\lambda \in \Lambda$, $\|\lambda\|_{L_2([0,T])} < L_2$.*
- *Or $T_n \to \infty$, $\varepsilon_n \equiv 1$ and there exist $L_1 < \infty$, $L_2 < \infty$ such that for all $\lambda \in \Lambda$, $\|\lambda\|_{L_1(\mathbb{R})} < L_1$, $\|\lambda\|_{L_2(\mathbb{R})} < L_2$.*

Then

$$\Delta(\mathscr{Q}_n, \mathscr{W}_n) \to 0 \quad \text{and} \quad \Delta(\mathscr{P}_n, \mathscr{Q}_n) \to 0 \text{ as } n \to \infty,$$

as soon as one of the following two conditions holds

1. *$\mathscr{G}$ is a subclass of discrete distributions with support on $\mathbb{Z}$: In this case an upper bound for the rate of convergence is $O\Big(\sqrt{\Delta_n} + T_n\Delta_n^{2\alpha}\varepsilon_n^{-2} + T_n\Delta_n\Big)$.*
2. *$\mathscr{G}$ is a subclass of absolutely continuous distributions with respect to the Lebesgue measure on $\mathbb{R}$ with uniformly bounded densities on a fixed neighborhood of 0: In this case an upper bound for the rate of convergence is $O\Big(\sqrt[4]{\Delta_n} + T_n\Delta_n^{2\alpha}\varepsilon_n^{-2} + T_n\Delta_n\Big)$.*

The paper is organized as follows. Sections 3.2.1 to 3.2.3 fix assumptions and notation. The main results, as well as examples, are given in Section 3.2.4. A discussion of the results can be found in Section 3.2.5. The proofs are postponed to Section 3.3. They are obtained as a sequence of results proving different (asymptotic) equivalences. Loosely speaking, we first reduce to having in each interval of the discretization at most one jump (Bernoulli approximation, Section 3.3.1). Secondly, we filter it out via an explicit Markov kernel, reducing ourselves to treating independent Gaussian variables (Section 3.3.2). Finally, we apply an argument similar to that in Brown, Low (1996) (Section 3.3.3) and collect all the pieces to conclude the proofs in Section 3.3.4. An appendix collects some proofs of general facts about the Le Cam distance that we use in the rest of the paper.

3.2 Assumptions and main results

3.2.1 Additive processes

Time inhomogeneous jump-diffusion processes are a special case of additive processes. Here we briefly recall definitions and properties of this class of processes.

Definition 3.2.1. A stochastic process $\{X_t\}_{t\geq 0}$ on $\mathbb{R}$ defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ is an *additive process* if the following conditions are satisfied.

1. $X_0 = 0$ $\mathbb{P}$-a.s.
2. Independent increments: for any choice of $n \geq 1$ and $0 \leq t_0 < t_1 < \ldots < t_n$, random variables X_{t_0}, $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.
3. There is $\Omega_0 \in \mathscr{A}$ with $\mathbb{P}(\Omega_0) = 1$ such that, for every $\omega \in \Omega_0$, $X_t(\omega)$ is right-continuous in $t \geq 0$ and has left limits in $t > 0$.
4. Stochastic continuity: $\forall \varepsilon > 0, \mathbb{P}(|X_{t+h} - X_t| \geq \varepsilon) \to 0$ as $h \to 0$.

Thanks to the *Lévy-Khintchine formula* (see Cont, Tankov (2004), Theorem 14.1), the characteristic function of any additive process $X = \{X_t\}_{t\in[0,T]}$ can be expressed, for all u in $\mathbb{R}$, as:

$$\mathbb{E}\big[e^{iuX_t}\big] = \exp\Big(iu\int_0^t f(r)dr - \frac{u^2}{2}\int_0^t \sigma^2(r)dr - \int_{\mathbb{R}}(1 - e^{iuy} + iuy\mathbb{I}_{|y|\leq 1})\nu_t(dy)\Big), \quad (3.3)$$

where $f(\cdot)$ and $\sigma^2(\cdot)$ belongs to $L_1(\mathbb{R})$ and ν_t is a positive measure on $\mathbb{R}$ satisfying

$$\nu_t(\{0\}) = 0 \text{ and } \int_{\mathbb{R}}(y^2 \wedge 1)\nu_t(dy) < \infty, \quad \forall t \in [0,T].$$

and a monotonicity condition: if $0 < s \leq t$ then $\nu_s(A) \leq \nu_t(A)$, for all $A \in \mathscr{B}(\mathbb{R})$ (see Sato 1999, Theorem 9.8). In the sequel we shall refer to $(f(t), \sigma^2(t), \nu_t)_{t\in[0,T]}$ as the local characteristics of the process X and a ν_t as above will be called a *Lévy measure*, for all t. This data characterizes uniquely the law of the process X. In the case where $f(\cdot)$ and $\sigma(\cdot)$ are constant functions and $\nu_t = \nu$ for all t, the process X satisfying (3.3) is stationary, and is called a *Lévy process* of characteristic triplet (f, σ^2, ν).

Let $D = D([0,\infty),\mathbb{R})$ be the space of mappings ω from $[0,\infty)$ into $\mathbb{R}$ that are right-continuous with left limits. Define the *canonical process* $x : D \to D$ by

$$\forall \omega \in D, \quad x_t(\omega) = \omega_t, \ \ \forall t \geq 0.$$

Let $\mathscr{D}_t$ and $\mathscr{D}$ be the σ-algebras generated by $\{x_s : 0 \leq s \leq t\}$ and $\{x_s : 0 \leq s < \infty\}$, respectively. Let X be an additive process defined on $(\Omega, \mathscr{A}, \mathbb{P})$ having local characteristics $(f(t), \sigma^2(t), \nu_t)_{t\in[0,T]}$. It is well known that it induces a probability measure $P^{(f,\sigma^2,\nu)}$ on $(D,\mathscr{D})$ such that $\{x_t\}$ defined on $\big(D,\mathscr{D},P^{(f,\sigma^2,\nu)}\big)$ is an additive process identical in law with $(\{X_t\},\mathbb{P})$ (that is the local characteristics of $\{x_t\}$ under $P^{(f,\sigma^2,\nu)}$ is $(f(t),\sigma^2(t),\nu_t)_{t\geq 0})$.

In the sequel we will denote by $\big(\{x_t\}, P^{(f,\sigma^2,\nu)}\big)$ such an additive process, stressing the probability measure and by $P_t^{(f,\sigma^2,\nu)}$ for the restriction of $P^{(f,\sigma^2,\nu)}$ to $\mathscr{D}_t$.

Further, for every function ω in D, we will denote by $\Delta\omega_r$ its jump at the time r and by ω^c, ω^d its continuous and discontinuous part, respectively:

$$\Delta\omega_r = \omega_r - \lim_{s\uparrow r} \omega_s, \ \ \omega_t^d = \sum_{r\leq t} \Delta\omega_r, \ \ \omega_t^c = \omega_t - \omega_t^d.$$

Note that, if $\nu_t = 0$ for all $t \geq 0$, then $\big(\{x_t\}, P^{(f,\sigma^2,0)}\big)$ is a Gaussian process that can be represented on $(\Omega,\mathscr{A},\mathbb{P})$ as

$$X_t = \int_0^t f(s)ds + \int_0^t \sigma(s)dW_s, \quad t \geq 0, \tag{3.4}$$

for some standard Brownian motion W on $(\Omega,\mathscr{A},\mathbb{P})$.

A time inhomogeneous jump-diffusion process as in (3.1), observed until the time T_n, is an additive process (apart from the possibly non-zero initial condition) with local characteristics $(f(t),\sigma^2(t),\nu_t)_{t\in[0,T_n]}$, where $\nu_t(\cdot) = \lambda(t)G(\cdot)$. We will write $\big(\{x_t\}, P_{T_n}^{(f,\sigma^2,\lambda G)}\big)$ for such a process. Also observe that $\big(\{x_t^c\}, P_{T_n}^{(f,\sigma^2,\lambda G)}\big)$ has the same law as $\big(\{x_t\}, P_{T_n}^{(f,\sigma^2,0)}\big)$. Moreover, thanks to the independence of the increments, the law of the i-th increment of (3.1) is the convolution product between the Gaussian law $\mathscr{N}\Big(\int_{t_{i-1}}^{t_i} f(s)ds, \int_{t_{i-1}}^{t_i}\sigma^2(s)ds\Big)$ and the law of the variable $\sum_{j=1}^{P_i} Y_j$, where P_i is Poisson of intensity $\lambda_i = \int_{t_{i-1}}^{t_i}\lambda(s)ds$.

3.2.2 Le Cam theory of statistical experiments

A *statistical model* is a triplet $\mathscr{P}_j = (\mathscr{X}_j, \mathscr{A}_j, \{P_{j,\theta}; \theta \in \Theta\})$ where $\{P_{j,\theta}; \theta\in\Theta\}$ is a family of probability distributions all defined on the same σ-field $\mathscr{A}_j$ over the *sample space* $\mathscr{X}_j$ and Θ is the *parameter space*. The *deficiency* $\delta(\mathscr{P}_1,\mathscr{P}_2)$ of $\mathscr{P}_1$ with respect to $\mathscr{P}_2$ quantifies “how much information we lose” by using $\mathscr{P}_1$ instead of $\mathscr{P}_2$ and is defined as $\delta(\mathscr{P}_1,\mathscr{P}_2)=\inf_K\sup_{\theta\in\Theta}||KP_{1,\theta}-P_{2,\theta}||_{TV}$, where TV stands for "total variation" and the infimum is taken over all "transitions" K (see Le Cam (1986), page 18). The general definition of transition is quite involved but, for our purposes, it is enough to know that Markov kernels are special cases of transitions.

The Le Cam Δ-distance is defined as the symetrization of δ and it defines a pseudo-metric. When $\Delta(\mathscr{P}_1,\mathscr{P}_2)=0$ the two statistical models are said to be equivalent. Two sequences of statistical models $(\mathscr{P}_1^n)_{n\in\mathbb{N}}$ and $(\mathscr{P}_2^n)_{n\in\mathbb{N}}$ are called *asymptotically equivalent* if $\Delta(\mathscr{P}_1^n,\mathscr{P}_2^n)$ tends to zero as n goes to infinity. There are various techniques to bound the Δ-distance. We report below only the properties that are useful for our purposes. For the proofs see, e.g., Le Cam (1986); Strasser (1985) and the Appendix.

Property 3.2.2. *Let $\mathscr{P}_j=(\mathscr{X},\mathscr{A},\{P_{j,\theta};\theta\in\Theta\})$, $j=1,2$, be two statistical models having the same sample space and define $\Delta_0(\mathscr{P}_1,\mathscr{P}_2):=\sup_{\theta\in\Theta}\|P_{1,\theta}-P_{2,\theta}\|_{TV}$. Then, $\Delta(\mathscr{P}_1,\mathscr{P}_2)\leq\Delta_0(\mathscr{P}_1,\mathscr{P}_2)$.*

In particular, Property 3.2.2 allows us to bound the Δ-distance between statistical models sharing the same sample space by means of classical bounds for the total variation distance. Classical bounds on the latter will thus prove useful:

Fact 3.2.3 (see Le Cam (1969), p. 35)**.** *Let P_1 and P_2 be two probability measures on $\mathscr{X}$, dominated by a common measure ξ, with densities $g_i=\frac{dP_i}{d\xi}$, $i=1,2$. Define*

$$L_1(P_1,P_2)=\int_{\mathscr{X}}|g_1(x)-g_2(x)|\xi(dx),$$
$$H(P_1,P_2)=\bigg(\int_{\mathscr{X}}\Big(\sqrt{g_1(x)}-\sqrt{g_2(x)}\Big)^2\xi(dx)\bigg)^{1/2}.$$

Then,

$$\frac{H^2(P,Q)}{2}\leq\|P_1-P_2\|_{TV}=\frac{1}{2}L_1(P_1,P_2)\leq H(P_1,P_2). \tag{3.5}$$

Fact 3.2.4. *[see Strasser (1985), Lemma 2.19] Let P and Q be two product measures defined on the same sample space: $P=\otimes_{i=1}^n P_i$, $Q=\otimes_{i=1}^n Q_i$. Then*

$$H^2(P,Q)\leq\sum_{i=1}^n H^2(P_i,Q_i). \tag{3.6}$$

Using (3.5)*, it follows that*

$$\|P-Q\|_{TV}\leq\sqrt{\sum_{i=1}^n 2\|P_i-Q_i\|_{TV}}.$$

Below, we collect some well-known facts that can be used to establish asymptotic equivalences. For the convenience of the reader, their proofs can be found in the Appendix.

Fact 3.2.5. *Let $Q_1 \sim \mathscr{N}(\mu_1, \sigma_1^2)$ and $Q_2 \sim \mathscr{N}(\mu_2, \sigma_2^2)$. Then*

$$\|Q_1 - Q_2\|_{TV} \leq \sqrt{\Big(1 - \frac{\sigma_1}{\sigma_2}\Big)^2 + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}} \leq \sqrt{\Big(1 - \frac{\sigma_1^2}{\sigma_2^2}\Big)^2 + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}}.$$

Fact 3.2.6. *Let $m_i(\cdot)$ and $\sigma(\cdot)$ be real functions such that $\int_{\mathbb{R}} \frac{m_i(s)^2}{\sigma(s)^2} ds < \infty$, $i = 1, 2$, with $\sigma(\cdot) > 0$. Then, with the same notation as in Section 3.2.1:*

$$L_1\big(P_t^{(m_1,\sigma^2,0)}, P_t^{(m_2,\sigma^2,0)}\big) = 2\Big(1 - 2\phi\Big(-\frac{D_t}{2}\Big)\Big), \quad \forall t > 0,$$

where ϕ denotes the cumulative distribution function of a Gaussian random variable $\mathscr{N}(0,1)$ and

$$D_t^2 = \int_0^t \frac{(m_1(s) - m_2(s))^2}{\sigma^2(s)} ds.$$

In particular, $L_1\big(P_t^{(m_1,\sigma^2,0)}, P_t^{(m_2,\sigma^2,0)}\big) = O(D_t)$.

Property 3.2.7. *Let $\mathscr{P}_i = (\mathscr{X}_i, \mathscr{A}_i, \{P_{i,\theta}, \theta \in \Theta\})$, $i = 1, 2$, be two statistical models. Let $S : \mathscr{X}_1 \to \mathscr{X}_2$ be a sufficient statistics such that the distribution of S under $P_{1,\theta}$ is equal to $P_{2,\theta}$. Then $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$.*

3.2.3 The parameter space

We now state the different kinds of assumptions on the non-parametric classes $\mathscr{F}$, Λ and $\mathscr{G}$ that will show up in the statements of the theorems:

(F1) Every $f \in \mathscr{F}$ is continuous and $\sup_{t\in\mathbb{R}}\{|f(t)| : f \in \mathscr{F}\} \leq B$, for some constant B.

(F2) Defining:

$$\bar{f}_n(t) = \begin{cases} f(t_i) & \text{if} \quad t_{i-1} \leq t < t_i, \quad i = 1, \dots, n; \\ f(T_n) & \text{if} \quad t = T_n; \end{cases} \tag{3.7}$$

we have

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \int_0^{T_n} \frac{(f(t) - \bar{f}_n(t))^2}{\sigma_n^2(t)} dt = 0. \tag{3.8}$$

(L1) Denoting by $\|\cdot\|_1$ the L_1 norm on $\mathbb{R}$, we require $\sup_{\lambda\in\Lambda} \|\lambda\|_1 \leq L_1$, for some constant L_1.

(L2) Denoting by $\|\cdot\|_2$ the L_2 norm on $\mathbb{R}$, we ask $\sup_{\lambda\in\Lambda}\|\lambda\|_2^2 \leq L_2$, for some constant L_2.

(G1) $\mathscr{G}$ is a subset of discrete distributions concentrated on $\mathbb{Z}$.

(G2) $\mathscr{G}$ is a subset of the set of absolutely continuous distributions with respect to the Lebesgue measure, $h = \frac{dG}{d\mathrm{Leb}}$. We ask that there are absolute constants $N_1, N_2 > 0$ such that $h \leq N_2$ Leb-a.e. on $[-\frac{1}{N_1}, \frac{1}{N_1}]$.

3.2.4 Main results and examples

Recall that models (3.1) and (3.2) depend on diffusion coefficients $\sigma_n(\cdot) = \varepsilon_n\sigma(\cdot)$, where ε_n is either 1 (if $T_n \to \infty$) or $\varepsilon_n \to 0$ (if $T_n = T$ finite). We will assume that $\sigma(\cdot)$ is absolutely continuous, strictly positive, and its logarithmic derivative is uniformly bounded: There exists a constant C_1 such that:

$$\left|\frac{d}{dt}\ln\sigma(t)\right| \leq C_1, \quad t \in \mathbb{R}. \tag{3.9}$$

Our main results are then:

Theorem 3.2.8. *Suppose that the parameter space $\mathscr{F}$ fulfills the assumptions (F1) and (F2) and let $\sigma(\cdot)$ satisfy* (3.9) *as above. If, in addition, Λ and $\mathscr{G}$ satisfy Assumptions (L2) and (G1), respectively, then, for n big enough, we have*

$$\Delta\big(\mathscr{P}_n, \mathscr{Q}_n\big) = \Delta\big(\mathscr{Q}_n, \mathscr{W}_n\big) \leq O\bigg(\sup_{f\in\mathscr{F}}\int_0^{T_n}\frac{(f(t)-\bar f_n(t))^2}{\sigma_n^2(t)}dt + T_n\Delta_n + \sqrt{\Delta_n}\bigg).$$

Here, the O depends only on the constants C_1 and L_2.

Theorem 3.2.9. *Suppose that the parameter space $\mathscr{F}$ fulfills the assumptions (F1) and (F2) and let $\sigma(\cdot)$ satisfy* (3.9) *as above. Suppose also there exist m_σ, M_σ such that $0 < m_\sigma \leq \sigma(\cdot) \leq M_\sigma < \infty$. Moreover suppose that Λ fulfills Assumptions (L1), (L2) and $\mathscr{G}$ fulfills Assumption (G2). Then, for n big enough, we have*

$$\Delta(\mathscr{P}_n, \mathscr{Q}_n) = \Delta(\mathscr{W}_n, \mathscr{Q}_n) \leq O\bigg(\sup_{f\in\mathscr{F}}\int_0^{T_n}\frac{(f(t)-\bar f_n(t))^2}{\sigma_n^2(t)}dt + T_n\Delta_n + \Delta_n^{\frac{1}{4}}\bigg).$$

Here the leading terms in the O depend on L_1, N_2 and M_σ only.

As a corollary, when $\mathscr{F}$ consists of uniformly bounded α-Hölder functions, one retrieves the rates of convergence stated in Theorem 3.1.1. We now give some examples of situations where our results can be applied.

Example 3.2.10. The sum of a diffusion process and an inhomogeneous Poisson process: This corresponds to setting $Y_1 \equiv 1$, so that $\mathscr{G}$ consists of the only Dirac mass in 1. Let $\sigma_n(\cdot) = \varepsilon_n \sigma(\cdot)$ satisfy (3.9) as above, and Λ satisfy Assumption (L2). If $\mathscr{F}$ is a class of α-Hölder, uniformly bounded functions on $\mathbb{R}$ with $\alpha \in (0,1]$, for n big enough, an application of Theorem 3.2.8 yields:

$$\Delta(\mathscr{P}_n, \mathscr{Q}_n) = \Delta(\mathscr{Q}_n, \mathscr{W}_n) = O\Big(\sqrt{\Delta_n} + T_n\Delta_n + T_n\Delta_n^{2\alpha}\varepsilon_n^{-2}\Big).$$

Example 3.2.11. The time inhomogeneous Merton model: This corresponds to $\mathscr{G}$ being a parametric class of Gaussian random distributions $\mathscr{N}(m, \Gamma^2)$, $\Gamma > 0$. Suppose that $\sigma(\cdot)$ is as in Example 3.2.10 and Λ satisfies Assumptions (L1) and (L2). Let $\mathscr{F}$ be a class of α-Hölder, uniformly bounded functions on $\mathbb{R}$ with $\alpha \in (0,1]$. Then, for n big enough, an application of Theorem 3.2.9 yields:

$$\Delta(\mathscr{P}_n, \mathscr{Q}_n) = \Delta(\mathscr{Q}_n, \mathscr{W}_n) = O\Big(\sqrt[4]{\Delta_n} + T_n\Delta_n + T_n\Delta_n^{2\alpha}\varepsilon_n^{-2}\Big).$$

3.2.5 Discussion

Remark 3.2.12. Hypotheses (F1), (F2) are modeled on those in Brown, Low (1996). They are satisfied, for example, by any class $\mathscr{F}$ of uniformly bounded α-Hölder functions, with α depending on the asymptotics of the data Δ_n, T_n, ε_n, as well as by uniformly bounded Sobolev $W^{\alpha,2}$ functions. Hypothesis (3.9) on $\sigma^2(\cdot)$ also appears in Brown, Low (1996). The non-parametric classes Λ and $\mathscr{G}$ were introduced to stress that the precise parameters λ, G chosen do not play any role in the proofs.

Remark 3.2.13. In the case where $\mathscr{G}$ satisfies Assumption (G1) (i.e. the Y_i's are discrete), the Markov kernel K in Lemma 3.3.2 does not depend on $\sigma(\cdot)$. Hence, combining our Theorem 3.2.8 with the one by Carter (2007) one can obtain the same equivalence result when $\sigma(\cdot)$ is an unknown nuisance parameter.

Remark 3.2.14. An important advantage of showing the asymptotic equivalence between statistical models is that it allows to transfer statistical inference procedures from one model to the other. This is done in such a way that the asymptotic risk remains the same, at least for bounded loss functions. When the proof of such an equivalence is constructive, one can provide a precise recipe for producing, from a sequence of procedures in one problem, an asymptotically equivalent sequence in the other one. Formally, let us consider two sequences of statistical models $\mathscr{P}_j^n = (\mathscr{X}_{j,n}, \mathscr{A}_{j,n}, \{P_{j,n,\theta}; \theta \in \Theta\})$ and a decision or action space $(A, \mathscr{A})$. Furthermore, for every n, let us denote by $\rho_{j,n}$ a possibly *randomized decision procedure* in $\mathscr{P}_j^n$, i.e. a Markov kernel $\rho_{j,n} : (\mathscr{X}_{j,n}, \mathscr{A}_{j,n}) \mapsto (A, \mathscr{A})$

and by $R(\mathscr{P}_{j,n}, \rho_{j,n}, L_n, \theta)$ the *risk* in the model $\mathscr{P}_{j,n}$ with respect to the decision rule $\rho_{j,n}$ and the loss function L_n. One says that the sequences of procedures $\rho_{1,n}$ and $\rho_{2,n}$ are *asymptotically equivalent* if for any sequence of bounded loss function L_n one has $\lim_{n\to\infty} \sup_{\theta\in\Theta} |R(\mathscr{P}_{1,n}, \rho_{1,n}, L_n, \theta) - R(\mathscr{P}_{2,n}, \rho_{2,n}, L_n, \theta)| = 0$.

In this paper there are essentially four statistical models that we prove to be mutually asymptotically equivalent: $\mathscr{P}_n$, $\mathscr{W}_n$, $\mathscr{Q}_n$ and $\tilde{\mathscr{Q}}_n$. The latter is associated with the observation of the increments $(y_{t_i} - y_{t_{i-1}})_{i=1}^n$ of (y_t) defined as in (3.2). The proofs of Theorems 3.2.8 and 3.2.9 allow us to use the knowledge of a sequence of procedures in $\mathscr{P}_n$, $\mathscr{W}_n$ or $\tilde{\mathscr{Q}}_n$ for producing one in $\mathscr{Q}_n$.

For example, suppose that $\mathscr{G}$ satisfies Assumption (G1) and let (δ_n) be a sequence of procedures in $\tilde{\mathscr{Q}}_n$. Define a sequence of procedures in $\mathscr{Q}_n$ as:

$$\gamma_n(z_0, \dots, z_n) := \delta_n\big(z_1 - z_0 - [z_1 - z_0], \dots, z_n - z_{n-1} - [z_n - z_{n-1}]\big), \quad z_0, \dots, z_n \in \mathbb{R},$$

where $[z]$ denotes the the closest integer to z. Then (γ_n) is asymptotically equivalent to (δ_n).

Remark that, up to this point, we did not use the knowledge of $\sigma^2(\cdot)$. In particular, if one disposes of a sequence of estimators of $f(\cdot)$ in $\tilde{\mathscr{Q}}_n$ an equivalent one can be deduced in $\mathscr{Q}_n$ also when $\sigma^2(\cdot)$ is unknown.

3.3 Proofs

3.3.1 Bernoulli approximation

Lemma 3.3.1. *Let $(N_i)_{i=1}^n, (P_i)_{i=1}^n$, $(Y_i)_{i=1}^n$ and $(\varepsilon_i)_{i=1}^n$ be samples of, respectively, Gaussian random variables $\mathscr{N}(m_i, \sigma_i^2)$, Poisson random variables $\mathscr{P}(\lambda_i)$, random variables with common distribution and Bernoulli random variables of parameters $\alpha_i := \lambda_i e^{-\lambda_i}$. Suppose that $(N_i)_{i=1}^n, (P_i)_{i=1}^n$, $(Y_i)_{i=1}^n$ and $(\varepsilon_i)_{i=1}^n$ are all defined on the same probability space $(\Omega, \mathscr{A}, \mathbb{P})$ and that they are all independent. Let us denote by Q_{N_i} (resp. $Q_{(Y_i,P_i)}$, $Q_{(Y_1,\varepsilon_i)}$) the law of N_i (resp. $\sum_{j=1}^{P_i} Y_j$, $\varepsilon_i Y_1$). Then*

$$\| \otimes_{i=1}^n Q_{N_i} * Q_{(Y_i,P_i)} - \otimes_{i=1}^n Q_{N_i} * Q_{(Y_i,\varepsilon_i)} \|_{TV} \leq 2\sqrt{\sum_{i=1}^n \lambda_i^2} \tag{3.10}$$

where the symbol $*$ denotes the product convolution between measures.

Proof. Observe that:

$$\begin{aligned}\|Q_{N_i} * Q_{(Y_i,P_i)} - Q_{N_i} * Q_{(Y_1,\varepsilon_i)}\|_{TV} &= \sup_{A\in\mathscr{B}(\mathbb{R})} \bigg| \sum_{k\geq 0} \mathbb{P}\Big(N_i + \sum_{j=1}^{k} Y_j \in A\Big) e^{-\lambda_i}\frac{\lambda_i^k}{k!} \\ &\qquad - (1-\alpha_i)\mathbb{P}(N_i\in A) - \alpha_i \mathbb{P}(N_i + Y_1 \in A)\bigg| \\ &= \sup_{A\in\mathscr{B}(\mathbb{R})} \bigg| \sum_{k\geq 2} \Big(\mathbb{P}\Big(N_i + \sum_{j=1}^{k} Y_j \in A\Big) - \mathbb{P}(N_i\in A)\Big) e^{-\lambda_i}\frac{\lambda_i^k}{k!}\bigg| \\ &\leq 2\sum_{k\geq 2} e^{-\lambda_i}\frac{\lambda_i^k}{k!} \leq 2\lambda_i^2.\end{aligned}$$

We get (3.10) thanks to Fact 3.2.4. □

3.3.2 Explicit construction of Markov kernels

Lemma 3.3.2. *Let $(N_i)_{i=1}^n$ and $(\varepsilon_i)_{i=1}^n$ be samples of, respectively, Gaussian random variables $\mathscr{N}(m_i,\sigma_i^2)$ with $|m_i|\leq \frac{1}{3}$ and Bernoulli random variables of parameters $\alpha_i := \lambda_i e^{-\lambda_i}$, $\lambda_i > 0$. Moreover, let Y_1 be a discrete random variable taking values in $\mathbb{Z}$ and denote by Q_{N_i} (resp. $Q_{(Y_1,\varepsilon_i)}$) the law of N_i (resp. $\varepsilon_i Y_1$). For all x in $\mathbb{R}$ denote by $[x]$ the nearest integer to x and define the Markov kernel*

$$K(x,A) = \mathbb{I}_A(x-[x]), \quad \forall A \in \mathscr{B}(\mathbb{R}).$$

Then

$$\| \otimes_{i=1}^n K(Q_{N_i} * Q_{(Y_1,\varepsilon_i)}) - \otimes_{i=1}^n Q_{N_i}\|_{TV} \leq \sqrt{2\sum_{i=1}^{n}\Big(\frac{6}{\sigma_i}\varphi\Big(\frac{1}{6\sigma_i}\Big) + 4\phi\Big(\frac{-1}{6\sigma_i}\Big)\Big)} \tag{3.11}$$

where $*$ stands for the convolution product, ϕ denotes the cumulative distribution of a Gaussian random variable $\mathscr{N}(0,1)$ and φ the derivative of ϕ.

Proof. Denote by $g_i(\cdot)$ the density of N_i, by $h(\cdot)$ the density of Y_1 with respect to the counting measure and define $G_i(x,k) := (1-\alpha_i)g_i(x) + \alpha_i g_i(x-k)$, $\forall x\in\mathbb{R}$, $\forall k \in \mathbb{Z}$. We have, for all i:

$$\begin{aligned}\|K(Q_{N_i}*Q_{(Y_1,\varepsilon_i)})-Q_{N_i}\|_{TV} &= \sup_{A\in\mathscr{B}(\mathbb{R})}\bigg|\int \mathbb{I}_A(x-[x])\Big[(1-\alpha_i)g_i(x)+\alpha_i\sum_{k\in\mathbb{Z}}h(k)g_i(x-k)\Big]dx\\ &\qquad -\int\mathbb{I}_A(x)g_i(x)dx\bigg|\\ &\leq \sup_{A\in\mathscr{B}(\mathbb{R})}\sum_{k\in\mathbb{Z}}h(k)\bigg|\int\Big(\mathbb{I}_A(x-[x])G_i(x,k)-\mathbb{I}_A(x)g_i(x)\Big)dx\bigg|.\end{aligned}$$

Writing $\int\mathbb{I}_A(x-[x])G_i(x,k)dx$ as $\sum_{l\in\mathbb{Z}}\int_{-\frac12}^{\frac12}\mathbb{I}_A(x)G_i(x+l,k)dx$, one can bound $\Big|\int\big(\mathbb{I}_A(x-[x])G_i(x,k)-\mathbb{I}_A(x)g_i(x)\big)dx\Big|$ by the sum of the following three terms:

$$\begin{aligned}I&=\bigg|\int_{-\frac12}^{\frac12}\mathbb{I}_A(x)\Big[G_i(x,k)+G_i(x+k,k)-g_i(x)\Big]dx\bigg|\\ &=\bigg|\int_{-\frac12}^{\frac12}\mathbb{I}_A(x)\big[\alpha_ig_i(x-k)+(1-\alpha_i)g_i(x+k)\big]dx\bigg|\leq\int_{-\frac12}^{\frac12}\big(g_i(x-k)+g_i(x+k)\big)dx\\ II&=\sum_{l\in\mathbb{Z}^*-\{k\}}\int_{-\frac12}^{\frac12}|G_i(x+l,k)|dx\leq\sum_{l\in\mathbb{Z}^*-\{k\}}\int_{-\frac12}^{\frac12}\big(g_i(x+l)+g_i(x+l-k)\big)dx\\ III&=\int_{[-\frac12,\frac12]^c}g_i(x)dx.\end{aligned}$$

Since $\Big|\int\big(\mathbb{I}_A(x-[x])G_i(x,0)-\mathbb{I}_A(x)g_i(x)\big)dx\Big|\leq\int_{[-\frac12,\frac12]^c}g_i(x)dx$ and $h(0)\leq 1$, we obtain

$$\begin{aligned}\|K(Q_{N_i}*Q_{(Y_1,\varepsilon_i)})-Q_{N_i}\|_{TV}&\leq\sum_{k\in\mathbb{Z}^*}h(k)\int_{-\frac12}^{\frac12}\big(g_i(x-k)+g_i(x+k)\big)dx\\ &\quad+\sum_{k\in\mathbb{Z}^*,l\in\mathbb{Z}^*-\{k\}}h(k)\int_{-\frac12}^{\frac12}\big(g_i(x+l)+g_i(x+l-k)\big)dx\\ &\quad+2\int_{[-\frac12,\frac12]^c}g_i(x)dx.\end{aligned}$$

Using the mean value theorem one can write

$$\begin{aligned}\int_{-\frac12}^{\frac12}\big(g_i(x-k)+g_i(x+k)\big)dx&=\phi\Big(\frac{1/2-k-m_i}{\sigma_i}\Big)-\phi\Big(\frac{-1/2-k-m_i}{\sigma_i}\Big)\\ &\quad+\phi\Big(\frac{1/2+k-m_i}{\sigma_i}\Big)-\phi\Big(\frac{-1/2+k-m_i}{\sigma_i}\Big)\\ &=\frac{1}{\sigma_i}\big(\varphi(\xi_{1,k})+\varphi(\xi_{2,k})\big)\end{aligned}$$

for some $\xi_{1,k} \in \left[\frac{-1/2-k-m_i}{\sigma_i}, \frac{1/2-k-m_i}{\sigma_i}\right]$ and $\xi_{2,k} \in \left[\frac{-1/2+k-m_i}{\sigma_i}, \frac{1/2+k-m_i}{\sigma_i}\right]$. In particular, since $|m_i| \leq \frac{1}{3}$ one has that $\varphi(\xi_{j,k}) \leq \varphi\Big(\frac{1}{6\sigma_i}\Big)$, $j = 1, 2$, hence

$$\sum_{k\in\mathbb{Z}^*} h(k) \int_{-\frac{1}{2}}^{\frac{1}{2}} \big(g_i(x-k) + g_i(x+k)\big)dx \leq \sum_{k\in\mathbb{Z}^*} \frac{2h(k)}{\sigma_i}\varphi\Big(\frac{1}{6\sigma_i}\Big) \leq \frac{2}{\sigma_i}\varphi\Big(\frac{1}{6\sigma_i}\Big).$$

In the same way one can write

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \big(g_i(x+l) + g_i(x+l-k)\big)dx = \frac{1}{\sigma_i}\big(\varphi(\eta_{1,l}) + \varphi(\eta_{2,l-k})\big)$$

for some $\eta_{1,l} \in \left[\frac{-1/2+l-m_i}{\sigma_i}, \frac{1/2+l-m_i}{\sigma_i}\right]$ and $\eta_{2,l-k} \in \left[\frac{-1/2+l-k-m_i}{\sigma_i}, \frac{1/2+l-k-m_i}{\sigma_i}\right]$. Then:

$$\begin{aligned}
\sum_{k\in\mathbb{Z}^*, l\in\mathbb{Z}^*-\{k\}} h(k) \int_{-\frac{1}{2}}^{\frac{1}{2}} \big(g_i(x+l) + g_i(x+l-k)\big)dx &\leq \sum_{k\in\mathbb{Z}^*, l\in\mathbb{Z}^*-\{k\}} \frac{h(k)}{\sigma_i}\big(\varphi(\eta_{1,l}) + \varphi(\eta_{2,l-k})\big) \\
&\leq \sum_{k\in\mathbb{Z}^*, l\in\mathbb{Z}^*-\{k\}} \frac{h(k)}{\sigma_i}\varphi(\eta_{1,l}) + \sum_{k,w\in\mathbb{Z}^*} \frac{h(k)}{\sigma_i}\varphi(\eta_{2,w}) \\
&\leq \sum_{k\in\mathbb{Z}^*} \frac{h(k)}{\sigma_i} \sum_{l\in\mathbb{Z}^*} \varphi(\eta_{1,l}) + \sum_{k\in\mathbb{Z}^*} \frac{h(k)}{\sigma_i} \sum_{w\in\mathbb{Z}^*} \varphi(\eta_{2,w}) \\
&\leq \frac{1}{\sigma_i} \sum_{l\in\mathbb{Z}^*} \big(\varphi(\eta_{1,l}) + \varphi(\eta_{2,l})\big).
\end{aligned}$$

Now, $|\eta_{i,l}| \geq \frac{|l|-5/6}{\sigma_i}$, $i = 1, 2$, so

$$\begin{aligned}
\frac{1}{\sigma_i} \sum_{l\in\mathbb{Z}^*} \big(\varphi(\eta_{1,l}) + \varphi(\eta_{2,l})\big) &\leq \frac{4}{\sigma_i}\varphi\Big(\frac{1}{6\sigma_i}\Big) + \frac{1}{\sigma_i}\sum_{|l|\geq 2} \varphi\Big(\frac{|l|-5/6}{\sigma_i}\Big) \\
&\leq \frac{4}{\sigma_i}\varphi\Big(\frac{1}{6\sigma_i}\Big) + 2\int_{\frac{1}{6\sigma_i}}^{\infty} \varphi(x)dx.
\end{aligned}$$

Finally, $\int_{[-\frac{1}{2},\frac{1}{2}]^c} g_i(x)dx \leq \int_{[-\frac{1}{6\sigma_i},\frac{1}{6\sigma_i}]^c} \varphi(x)dx = 2\phi\Big(-\frac{1}{6\sigma_i}\Big)$. Using Fact 3.2.4, these computations imply (3.11). □

Remark 3.3.3. In the case where $Y_1 \equiv 1$ (see Example 3.2.10) one can also consider a, maybe, more natural Markov kernel, that is:

$$K(x, A) = \mathbb{I}_A(\Psi(x)), \quad \text{with} \quad \Psi(x) = \begin{cases} x & \text{if } x \leq \frac{1}{2}, \\ x - 1 & \text{otherwise.} \end{cases}$$

However, the rate of convergence in (3.11) turns out to be asymptotically the same regardless of the chosen kernel.

Lemma 3.3.4. *Let $(N_i)_{i=1}^n$ and $(\varepsilon_i)_{i=1}^n$ be samples of, respectively, Gaussian random variables $\mathscr{N}(m_i,\sigma_i^2)$ with $|m_i|\leq L$ for some constant L and Bernoulli random variables of parameters $\alpha_i := \lambda_i e^{-\lambda_i}$. Moreover, let Y_1 be a random variable with density $h(\cdot)$ with respect to the Lebesgue measure and denote by Q_{N_i} (resp. $Q_{(Y_1,\varepsilon_i)}$) the law of N_i (resp. $\varepsilon_i Y_1$). Fix a $0<\varepsilon<1$ and define, for all i, the Markov kernel*

$$K_i(x,A)=\begin{cases}\mathbb{I}_A(x) & \text{if } x\in B_i := [-(L+\sigma_i^{1-\varepsilon}), L+\sigma_i^{1-\varepsilon}],\\ \frac{1}{\sqrt{2\pi\sigma_i^2}}\int_A e^{-\frac{y^2}{2\sigma_i^2}}dy, & \text{if } x\in B_i^c.\end{cases}$$

Then

$$\big\|\otimes_{i=1}^n K_i(Q_{N_i}*Q_{(Y_1,\varepsilon_i)})-\otimes_{i=1}^n Q_{N_i}\big\|_{TV}\leq\sqrt{2\sum_{i=1}^n\bigg(8\phi(-\sigma_i^{-\varepsilon})+\frac{\alpha_i|m_i|}{\sqrt{2}\sigma_i}+2\alpha_i\int_{-2\beta_i}^{2\beta_i}h(y)dy\bigg)}$$

where ϕ denotes the cumulative distribution of a Gaussian random variable $\mathscr{N}(0,1)$ and $\beta_i = L+\sigma_i^{1-\varepsilon}$.

Proof. The total variation distance between the measures $K_i(Q_{N_i}*Q_{(Y_1,\varepsilon_i)})$ and Q_{N_i} is bounded by the sum of the following two terms:

$$I=\sup_{A\in\mathscr{B}(\mathbb{R})}|K_i(Q_{N_i}*Q_{(Y_1,\varepsilon_i)})(A\cap B_i)-Q_{N_i}(A\cap B_i)|,$$
$$II=\sup_{A\in\mathscr{B}(\mathbb{R})}|K_i(Q_{N_i}*Q_{(Y_1,\varepsilon_i)})(A\cap B_i^c)-Q_{N_i}(A\cap B_i^c)|.$$

Denote by $Q_{\tilde{N}_i}$ the distribution of the Gaussian random variable $\tilde{N}_i \sim \mathscr{N}(0,\sigma_i^2)$, then

$$\begin{aligned} I &= \sup_{A\in\mathscr{B}(\mathbb{R})}\bigg|\alpha_i\Big(\mathbb{P}(N_i+Y_1\in A\cap B_i)+\mathbb{P}(\tilde{N}\in A\cap B_i)\mathbb{P}(N_i+Y_1\in B_i^c)\Big) \\ &\qquad\qquad + (1-\alpha_i)\mathbb{P}(\tilde{N}_i\in A\cap B_i)\mathbb{P}(N_i\in B_i^c)-\alpha_i\mathbb{P}(N_i\in A\cap B_i)\bigg| \\ &\le \sup_{A\in\mathscr{B}(\mathbb{R})}\alpha_i\bigg(\mathbb{P}\big(N_i+Y_1\in A\cap B_i\big)+\Big|\mathbb{P}(N_i\in A\cap B_i)\Big[\mathbb{P}(N_i+Y_1\in B_i^c)-1\Big]\Big| \\ &\qquad\qquad + \Big|\mathbb{P}(N_i+Y_1\in B_i^c)\Big[\mathbb{P}(\tilde{N}\in A\cap B_i)-\mathbb{P}(\tilde{N}\in A\cap B_i)\Big]\Big|\bigg)+\mathbb{P}(N_i\in B_i^c) \\ &\le \alpha_i\big(2\mathbb{P}(N_i+Y_1\in B_i)+\|Q_{\tilde{N}_i}-Q_{N_i}\|_{TV}\big)+\mathbb{P}(N_i\in B_i^c) \end{aligned}$$

and

$$\begin{aligned} II &= \sup_{A\in\mathscr{B}(\mathbb{R})}\big|\mathbb{P}(\tilde{N}_i\in A\cap B_i^c)\mathbb{P}(N_i+\varepsilon_iY_1\in B_i^c)-\mathbb{P}(N_i\in A\cap B_i^c)\big| \\ &\le \mathbb{P}(\tilde{N}_i\in B_i^c)+\mathbb{P}(N_i\in B_i^c). \end{aligned}$$

Now observe that

- $\mathbb{P}(N_i+Y_1\in B_i)\le\mathbb{P}(|Y_1|>2\beta_i)\mathbb{P}(|N_i|>\beta_i)+\mathbb{P}(|Y_1|\le 2\beta_i)\le\mathbb{P}(N_i\in B_i^c)+\int_{-2\beta_i}^{2\beta_i}h(y)dy$,
- $\mathbb{P}(N_i\in B_i^c)=\phi\Big(-\frac{L+\sigma_i^{1-\varepsilon}+m_i}{\sigma_i}\Big)+1-\phi\Big(\frac{L+\sigma_i^{1-\varepsilon}-m_i}{\sigma_i}\Big)\le\phi(-\sigma_i^{-\varepsilon})+1-\phi(\sigma_i^{-\varepsilon})=2\phi(-\sigma_i^{-\varepsilon})$
- $\mathbb{P}(\tilde{N}_i\in B_i^c)=\phi\Big(-\frac{L+\sigma_i^{1-\varepsilon}}{\sigma_i}\Big)+1-\phi\Big(\frac{L+\sigma_i^{1-\varepsilon}}{\sigma_i}\Big)\le 2\phi(-\sigma_i^{-\varepsilon})$

Combining these bounds with Fact 3.2.5 we get:

$$I+II\le 8\phi(\sigma_i^{-\varepsilon})+\alpha_i\int_{-2\beta_i}^{2\beta_i}h(y)dy+\alpha_i\frac{|m_i|}{\sqrt{2}\sigma_i}.$$

An application of Fact 3.2.4 allows us to conclude the proof.

□

3.3.3 Asymptotic equivalence between discretely and continuously observed Gaussian processes

Recall that $\tilde{\mathscr{Q}}_n$ is the statistical model associated with the observation of the increments $(y_{t_i}-y_{t_{i-1}})_{i=1}^n$ of (y_t) defined as in (3.2). Then we have:

Proposition 3.3.5. *Suppose that the parameter space $\mathscr{F}$ fulfills Assumption (F2) and let $\sigma(\cdot) > 0$ be a given absolutely continuous functions on $\mathbb{R}$ satisfying* (3.9)*. Then, the statistical models $\mathscr{W}_n$ and $\tilde{\mathscr{Q}}_n$ are asymptotically equivalent as n goes to infinity. An upper bound for the rate of convergence is given by $O\Big(\sup_{f\in\mathscr{F}}\int_0^{T_n}\frac{(f(s)-\bar{f}_n(s))^2}{\sigma_n^2(s)}ds + T_n\Delta_n\Big)$.*

Proof. The proof is based on the same ideas as in Brown, Low (1996). However, since some modifications are needed, we include a complete proof for the convenience of the reader.

STEP 1: We start by considering the statistical model, $\bar{\mathscr{P}}_n$, associated with a Gaussian process on $[0, T_n]$ with local characteristic $(\bar{f}_n(t), \sigma_n^2(t), 0)_{t\in[0,T_n]}$ (see (3.7) for the definition of $\bar{f}_n(\cdot)$). Fact 3.2.6 guarantees that

$$\Delta(\mathscr{P}_n, \bar{\mathscr{P}}_n) = O\Big(\sup_{f\in\mathscr{F}}\int_0^{T_n}\frac{(f(s)-\bar{f}_n(s))^2}{\sigma_n^2(s)}ds\Big).$$

STEP 2: By means of the Fisher factorization theorem, one can easily prove that the statistic defined by

$$S(\omega) = \bigg(\int_0^{t_1}\frac{d\omega_t}{\sigma_n^2(t)}, \dots, \int_{t_{n-1}}^{T_n}\frac{d\omega_t}{\sigma_n^2(t)}\bigg)$$

is a sufficient statistic for the family of probabilities $\{P_{T_n}^{(\bar{f}_n,\sigma_n^2,0)} : f \in \mathscr{F}\}$. Moreover, the law of S under $P_{T_n}^{(\bar{f}_n,\sigma_n^2,0)}$ is the law on $\mathbb{R}^n$ of a vector composed by n independent Gaussian random variable $\mu_i := \mathscr{N}\Big(f(t_i)\int_{t_{i-1}}^{t_i}\frac{dt}{\sigma_n^2(t)}, \int_{t_{i-1}}^{t_i}\frac{dt}{\sigma_n^2(t)}\Big)$, $i = 1, \dots, n$. Let us denote by $P_{i,f}$ the law on $\mathbb{R}$ of μ_i and by $\mathscr{S}_n$ the statistical model associated with the statistic S, that is

$$\mathscr{S}_n = \big\{\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), (\otimes_{i=1}^n P_{i,f} : f \in \mathscr{F})\big\}.$$

Then, by using Property 3.2.7, we get $\Delta(\bar{\mathscr{P}}_n, \mathscr{S}_n) = 0$. An application of the mean value theorem yields

$$\int_{t_{i-1}}^{t_i}\frac{ds}{\sigma_n^2(s)} = \frac{(t_i - t_{i-1})}{\sigma_n^2(\xi_i)}, \qquad \text{for a certain } \xi_i \text{ in } [t_{i-1}, t_i].$$

This allows us to pass from the model $\mathscr{S}_n$ to the equivalent one

$$\tilde{\mathscr{S}}_n = \big\{\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), (\otimes_{i=1}^n \tilde{P}_{i,f} : f \in \mathscr{F})\big\},$$

with $\tilde{P}_{i,f}$ denoting the distribution of a Gaussian random variable $\mathscr{N}(f(t_i)(t_i - t_{i-1}), \sigma_n^2(\xi_i)(t_i - t_{i-1}))$.

STEP 3: The last step consists in bounding the Δ-distance between $\tilde{\mathscr{S}}_n$ and $\tilde{\mathscr{Q}}_n$.

Property 3.2.2 and Facts 3.2.4–3.2.5 yield:

$$\Delta(\tilde{\mathscr{S}}_n, \tilde{\mathscr{Q}}_n) \leq \sup_{f\in\mathscr{F}} \sum_{i=1}^n \left[\left(1 - \frac{\sigma_n^2(\xi_i)(t_i - t_{i-1})}{\int_{t_{i-1}}^{t_i} \sigma_n^2(s)ds}\right)^2 + \frac{\left(\int_{t_{i-1}}^{t_i} \Big(f(t_i) - f(s)\Big)ds\right)^2}{2\int_{t_{i-1}}^{t_i} \sigma_n^2(s)ds} \right].$$

For all $i = 1, \dots, n$, let η_i and γ_i be elements in $[t_{i-1}, t_i]$ such that:

$$\int_{t_{i-1}}^{t_i} \sigma^2(s)ds = \sigma^2(\eta_i)(t_i - t_{i-1}), \quad \int_{t_{i-1}}^{t_i} f(s)ds = f(\gamma_i)(t_i - t_{i-1}).$$

By means of a Taylor expansion of $\sigma_n(\xi_i)/\sigma_n(\eta_i)$ we obtain

$$\frac{\sigma_n(\xi_i)}{\sigma_n(\eta_i)} = 1 + \frac{\sigma_n'(\eta_i)}{\sigma_n(\eta_i)}(\xi_i - \eta_i) + O(\xi_i - \eta_i)^2;$$

hence, thanks to assumption (3.9), we have

$$\left|\frac{\sigma_n(\xi_i)}{\sigma_n(\eta_i)}\right| \leq 1 + C_1(t_i - t_{i-1}) + O\big((t_i - t_{i-1})^2\big).$$

This means that

$$\Delta\big(\mathscr{S}_n, \tilde{\mathscr{Q}}_n\big) \leq \sup_{f\in\mathscr{F}} \sum_{i=1}^n \frac{(f(t_i) - f(\gamma_i))^2}{2\sigma_n^2(\eta_i)}(t_i - t_{i-1}) + O(T_n\Delta_n).$$

Here, the constant C_1 is hidden in the O. Observe that $\sum_{i=1}^n \frac{(f(t_i)-f(\gamma_i))^2}{2\sigma_n^2(\eta_i)}(t_i - t_{i-1})$ is less than $\int_0^{T_n} \frac{(f(s)-\bar{f}_n(s))^2}{2\sigma_n^2(s)}ds$. Indeed, on the one hand one can write:

$$\frac{(f(\xi_i) - f(t_i))^2}{\sigma_n^2(\eta_i)} = \frac{\left(\int_{t_{i-1}}^{t_i} \big(f(s) - f(t_i)\big)ds\right)^2}{(t_i - t_{i-1})\int_{t_{i-1}}^{t_i} \sigma_n^2(s)ds},$$

on the other hand, by means of the Hölder inequality, one has

$$\left(\int_{t_{i-1}}^{t_i} \big(f(s) - f(t_i)\big)ds\right)^2 \leq \int_{t_{i-1}}^{t_i} \sigma_n^2(s)ds \int_{t_{i-1}}^{t_i} \frac{\big(f(s) - f(t_i)\big)^2}{\sigma_n^2(s)}ds.$$

Combining these expressions one finds

$$\sum_{i=1}^n \frac{(f(t_i) - f(\gamma_i))^2}{2\sigma_n^2(\eta_i)} \leq \sum_{i=1}^n \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} \frac{(f(s) - f(t_i))^2}{2\sigma_n^2(s)}ds,$$

as claimed. □

Proposition 3.3.6. *Suppose that for every $f \in \mathscr{F}$, $\int_0^{T_n} \frac{f^2(s)}{\sigma_n^2(s)} ds < \infty$. Then, the statistical models $\mathscr{P}_n$ and $\mathscr{W}_n$ are equivalent.*

Proof. The Girsanov theorem assures that the measure $P^{(0,\sigma_n^2,\lambda G)}$ dominates the measure $P^{(f,\sigma_n^2,\lambda G)}$ and the density is given by

$$\frac{dP^{(f,\sigma_n^2,\lambda G)}}{dP^{(0,\sigma_n^2,\lambda G)}}(x) = \exp\bigg(\int_0^{T_n} \frac{f(s)}{\sigma_n^2(s)} dx_s^c - \frac{1}{2}\int_0^{T_n} \frac{f^2(s)}{\sigma_n^2(s)} ds\bigg).$$

We conclude the proof using Fact 3.2.7 applied to the statistic $S : \omega \to \omega^c$. $\square$

3.3.4 Proofs of Theorems 3.2.8 and 3.2.9

In order to prove our results we need to introduce the following notations:

$$m_i = \int_{t_{i-1}}^{t_i} f(s)ds, \quad \sigma_i^2 = \int_{t_{i-1}}^{t_i} \sigma_n^2(s)ds, \quad \lambda_i = \int_{t_{i-1}}^{t_i} \lambda(s)ds, \quad \alpha_i = \lambda_i e^{-\lambda_i}, \quad i = 1, \dots, n.$$

As a preliminary remark observe that the model $\mathscr{Q}_n$ is equivalent to the statistical model that observes the n increments $X_{t_i} - X_{t_{i-1}}$ of (3.1). Let us denote by $\hat{\mathscr{Q}}_n$ this latter and recall that the law of $X_{t_i} - X_{t_{i-1}}$ is the convolution product between the Gaussian law $\mathscr{N}\big(m_i, \sigma_i^2\big)$ and the law of the variable $\sum_{j=1}^{P_i} Y_j$, where P_i is Poisson with intensity λ_i. Regardless of the continuous or discrete nature of Y_1, the previous remark and Proposition 3.3.6 allow us to state that $\Delta(\mathscr{P}_n, \mathscr{Q}_n) = \Delta(\mathscr{P}_n, \hat{\mathscr{Q}}_n) = \Delta(\mathscr{W}_n, \hat{\mathscr{Q}}_n) = \Delta(\mathscr{Q}_n, \mathscr{W}_n)$. Now, to control $\Delta(\mathscr{P}_n, \hat{\mathscr{Q}}_n)$ suppose first that $\mathscr{G}$ satisfies Assumption (G1). On the one hand, for n big enough, $|m_i| \leq B(t_i - t_{i-1}) \leq \frac{1}{3}$, hence we can apply Lemmas 3.3.1–3.3.2 obtaining the bound:

$$\Delta\big(\hat{\mathscr{Q}}_n, \tilde{\mathscr{Q}}_n\big) \leq 2\sqrt{\sum_{i=1}^n \lambda_i^2} + \sqrt{2\sum_{i=1}^n \bigg(\frac{6}{\sigma_i}\varphi\Big(\frac{1}{6\sigma_i}\Big) + 4\phi\Big(\frac{-1}{6\sigma_i}\Big)\bigg)}.$$

Here we have implicitly used the following fact:

Let P_i be a probability measure on $(E_i, \mathcal{E}_i)$ and K_i a Markov kernel on $(G_i, \mathcal{G}_i)$. One can then define a Markov kernel K on $(\prod_{i=1}^n E_i, \otimes_{i=1}^n \mathcal{G}_i)$ such that $K(\otimes_{i=1}^n P_i) = \otimes_{i=1}^n K_i P_i$:

$$K(x_1, \dots, x_n; A_1 \times \dots \times A_n) := \prod_{i=1}^n K_i(x_i, A_i), \quad \forall x_i \in E_i,\ \forall A_i \in \mathcal{G}_i.$$

Also, observe that

$$2\sqrt{\sum_{i=1}^n \lambda_i^2} + \sqrt{2\sum_{i=1}^n \bigg(\frac{6}{\sigma_i}\varphi\Big(\frac{1}{6\sigma_i}\Big) + 4\phi\Big(\frac{-1}{6\sigma_i}\Big)\bigg)} = O\big(\sqrt{\Delta_n}\big),$$

where, in the leading term of the O, a constant L_2 is hidden. On the other hand, thanks to Proposition 3.3.5 we have:

$$\Delta(\tilde{\mathscr{Q}}_n, \mathscr{W}_n) \leq O\bigg(\sup_{f\in\mathscr{F}} \int_0^{T_n} \frac{(f(t)-\bar{f}_n(t))^2}{\sigma_n^2(t)} dt + T_n\Delta_n\bigg).$$

We then obtain the inequality stated in Theorem 3.2.8 by means of the triangular inequality.

In the same way, using Lemmas 3.3.1, 3.3.4 (taking $\varepsilon = \frac{1}{2}$) and Proposition 3.3.5 one can show the inequality in Theorem 3.2.9. Remark that one can actually choose any $0 < \varepsilon \leq \frac{1}{2}$. Smaller values of ε give better bounds for the term involving β_i in Theorem 3.2.9, but, if $\varepsilon \leq \frac{1}{2}$, under the hypotheses of Theorem 3.1.1, the leading term is $\sum_i \frac{\alpha_i B(t_i - t_{i_1})}{\sqrt{2}\sigma_i} = O\big(\sqrt{\Delta_n}\big)$ (here we hide the constants L_1, B and m_σ).

Appendix

Proofs of certain properties of the Le Cam Δ-distance

Proof of Fact 3.2.5. By symmetry, we can suppose $\sigma_1 \geq \sigma_2$. Denoting by g_i the density of Q_i with respect to Lebesgue, we have:

$$\frac{g_1}{g_2}(x) = \frac{\sigma_2}{\sigma_1} \exp\bigg(\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}\bigg).$$

Thus the Kullback-Leibler divergence is

$$\begin{aligned} D(Q_1, Q_2) = \int_{\mathbb{R}} g_1(x) \ln \frac{g_1(x)}{g_2(x)} dx &= \ln\frac{\sigma_2}{\sigma_1} + \int_{\mathbb{R}} \bigg(\frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{(x-\mu_1)^2}{2\sigma_1^2}\bigg) g_1(x) dx \\ &= \ln\frac{\sigma_2}{\sigma_1} + \frac{1}{2}\Big(\frac{\sigma_1^2}{\sigma_2^2} - 1\Big) + \frac{(\mu_1-\mu_2)^2}{2\sigma_1^2}. \end{aligned}$$

Let $r = \frac{\sigma_1}{\sigma_2} \geq 1$ and observe that

$$-\ln r + \frac{1}{2}(r^2 - 1) \leq (r-1)^2.$$

It is well known (see, e.g. Lemma 2.4 in Tsybakov (2009)) that the total variation distance is bounded by the square root of the Kullback-Leibler divergence, in this way we obtain:

$$\|Q_1 - Q_2\|_{TV} \leq \sqrt{\bigg(1 - \frac{\sigma_1}{\sigma_2}\bigg)^2 + \frac{(\mu_1-\mu_2)^2}{2\sigma_2^2}}.$$

□

Lemma 3.3.7. *Let g_i, $i = 1,2$ be the density of a Gaussian random variable $\mathscr{N}(\mu_i, \sigma^2)$. Then,*

$$L_1(g_1, g_2) = \mathbb{E}\bigg|\exp\left(X - \frac{(\mu_2-\mu_1)^2}{2\sigma^2}\right) - 1\bigg| = 2\Big[1 - 2\phi\Big(\frac{\mu_2-\mu_1}{2\sigma}\Big)\Big] \tag{3.12}$$

where $X \sim \mathcal{N}\left(0, \frac{(\mu_2-\mu_1)^2}{2\sigma^2}\right)$ and ϕ is the cumulative distribution function of a Gaussian random variable $\mathcal{N}(0,1)$.

Proof. Without loss of generality let us suppose that $\mu_1 \leq \mu_2$. Then we can write:

$$L_1(g_1,g_2) = \int_{\mathbb{R}} |g_1(x) - g_2(x)|dx = \int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} (g_1(x) - g_2(x))dx + \int_{\frac{\mu_1+\mu_2}{2}}^{\infty} (g_2(x) - g_1(x))dx.$$

Observe that

$$\int_{-\infty}^{\frac{\mu_1+\mu_2}{2}} g_1(x)dx = \mathbb{P}\left(\mathcal{N}(\mu_1,\sigma^2) \leq \frac{\mu_1+\mu_2}{2}\right) = \mathbb{P}\left(\mathcal{N}(0,1) \leq \frac{\mu_2-\mu_1}{2\sigma}\right) = \phi\left(\frac{\mu_2-\mu_1}{2\sigma}\right).$$

Similarly one has

$$\begin{aligned}\int_{\mathbb{R}} |g_1(x) - g_2(x)|dx &= \phi\left(\frac{\mu_2-\mu_1}{2\sigma}\right) - \phi\left(\frac{\mu_1-\mu_2}{2\sigma}\right) \\ &\quad + \left(1 - \phi\left(\frac{\mu_2-\mu_1}{2\sigma}\right)\right) - \left(1 - \phi\left(\frac{\mu_1-\mu_2}{2\sigma}\right)\right) \\ &= 2\left[\phi\left(\frac{\mu_1-\mu_2}{2\sigma}\right) - \phi\left(\frac{\mu_2-\mu_1}{2\sigma}\right)\right] = 2\left[1 - 2\phi\left(\frac{\mu_2-\mu_1}{2\sigma}\right)\right],\end{aligned}$$

thus,

$$L_1(g_1,g_2) = 2\Big[1 - 2\phi\Big(\frac{\mu_2-\mu_1}{2\sigma}\Big)\Big].$$

On the other hand we can also express the L_1-norm between g_1 and g_2 as

$$
\begin{aligned}
L_1(g_1,g_2) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}}\left|\exp\Big(-\frac{(x-\mu_1)^2}{2\sigma^2}\Big)-\exp\Big(-\frac{(x-\mu_2)^2}{2\sigma^2}\Big)\right|dx\\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}}\left|\exp\Big(-\frac{(x-\mu_2)^2-2(\mu_1-\mu_2)(x-\mu_2)+(\mu_1-\mu_2)^2}{2\sigma^2}\Big)\right.\\
&\qquad\qquad\left.-\exp\Big(-\frac{(x-\mu_2)^2}{2\sigma^2}\Big)\right|dx\\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{\mathbb{R}}\left|\exp\Big(\frac{2(\mu_1-\mu_2)(x-\mu_2)-(\mu_1-\mu_2)^2}{2\sigma^2}\Big)-1\right|\exp\Big(-\frac{(x-\mu_2)^2}{2\sigma^2}\Big)dx\\
&= \mathbb{E}\left|\exp\Big(\frac{2(\mu_1-\mu_2)(Y-\mu_2)-(\mu_1-\mu_2)^2}{2\sigma^2}\Big)-1\right|\\
&= \mathbb{E}\left|\exp\Big(\frac{(\mu_1-\mu_2)Z}{\sigma}-\frac{(\mu_1-\mu_2)^2}{2\sigma^2}\Big)-1\right|\\
&= \mathbb{E}\left|\exp\Big(X-\frac{(\mu_1-\mu_2)^2}{2\sigma^2}\Big)-1\right|,
\end{aligned}
$$

where $Y \sim \mathcal{N}(\mu_2,\sigma^2)$ et $Z \sim \mathcal{N}(0,1)$. □

Proof of Fact 3.2.6. Thanks to the Girsanov theorem one has that, $\forall\omega \in C$ and $\forall t > 0$

$$
\frac{dP_t^{(m_i,\sigma^2,0)}}{dP_t^{(0,\sigma^2,0)}}(\omega) = \exp\Big(\int_0^t \frac{m_i(t)}{\sigma^2(t)}d\omega_t - \frac{1}{2}\int_I \frac{m_i^2(t)}{\sigma^2(t)}dt\Big)P^{(0,\sigma^2,0)}(d\omega)
$$

In particular, $P_t^{(m_1,\sigma^2,0)}$ is absolutely continuous with respect to $P_t^{(m_2,\sigma^2,0)}$ and the density $g = \dfrac{dP_t^{(m_1,\sigma^2,0)}}{dP_t^{(m_2,\sigma^2,0)}}$ is given by:

$$
\begin{aligned}
g(\omega) &= \exp\Big(\int_0^t \frac{m_1(t)-m_2(t)}{\sigma^2(t)}d\omega_t - \frac{1}{2}\int_0^t \frac{m_1^2(t)-m_2^2(t)}{\sigma^2(t)}dt\Big)\\
&= \exp\Big(\int_0^t \frac{m_1(t)-m_2(t)}{\sigma^2(t)}(d\omega_t - m_2(t)dt) - \frac{1}{2}\int_0^t \frac{(m_1(t)-m_2(t))^2}{\sigma^2(t)}dt\Big).
\end{aligned}
\tag{3.13}
$$

Let us denote by $(Z_t)_{t\geq 0}$ the stochastic process satisfying the following EDS:

$$
dZ_t = m_2(t)dt + \sigma(t)dW_t, \quad t \geq 0,
$$

with $(W_t)_{t\geq 0}$ a standard Brownian motion. Then we have:,

$$L_1(P_t^{(m_1,\sigma^2,0)}, P_t^{(m_2,\sigma^2,0)}) = \int \big|g(\omega)-1\big|\frac{dP_t^{(m_2,\sigma^2,0)}}{dP_t^{(0,\sigma^2,0)}}(\omega)P^{(0,\sigma^2,0)}(d\omega)$$

$$= \mathbb{E}_{\mathbb{P}}\bigg|\exp\bigg(\int\frac{m_1(t)-m_2(t)}{\sigma^2(t)}(dZ_t-m_2(t)dt)-\frac{1}{2}\int\frac{(m_1(t)-m_2(t))^2}{\sigma^2(t)}dt\bigg)-1\bigg|$$

$$= \mathbb{E}_{\mathbb{P}}\bigg|\exp\bigg(\int\frac{(m_1(t)-m_2(t))}{\sigma^2(t)}\sigma(t)dW_t-\frac{1}{2}\int\frac{(m_1(t)-m_2(t))^2}{\sigma^2(t)}dt\bigg)-1\bigg|.$$

Observe that the random variable $\int_0^t\frac{(m_1(s)-m_2(s))}{\sigma(s)}dW_s$ has a centered Gaussian distribution with variance $\int_0^t\frac{(\mu(s)-\nu(s))^2}{\sigma^2(s)}ds$, thus, by means of Lemma 3.3.7, we can conclude that

$$L_1\Big(P_t^{(m_1,\sigma^2,0)}, P_t^{(m_2,\sigma^2,0)}\Big) = 2\bigg[1-2\phi\bigg(\frac{1}{2}\sqrt{\int_0^t\frac{(m_1(s)-m_2(s))^2}{\sigma^2(s)}ds}\bigg)\bigg].$$

□

Proof of Fact 3.2.7. In order to prove that $\delta(\mathscr{P}_1,\mathscr{P}_2)=0$ it is enough to consider the Markov kernel $M:(\mathscr{X}_1,\mathscr{A}_1)\to(\mathscr{X}_2,\mathscr{A}_2)$ defined as $M(x,B):=\mathbb{I}_B(S(x))$ $\forall x\in\mathscr{X}_1$ and $\forall B\in\mathscr{A}_2$. Conversely, to show that $\delta(\mathscr{P}_2,\mathscr{P}_1)=0$ one can consider the Markov kernel $K:(\mathscr{X}_2,\mathscr{A}_2)\to(\mathscr{X}_1,\mathscr{A}_1)$ defined as $K(y,A)=\mathbb{E}_{P_{2,\theta}}(\mathbb{I}_A|S=y)$, $\forall A\in\mathscr{A}_1$. Since S is a sufficient statistics, the Markov kernel K does not depend on θ. Denoting by $S_\# P_{1,\theta}$ the distribution of S under $P_{1,\theta}$, one has:

$$KP_{2,\theta}(A)=\int K(y,A)P_{2,\theta}(dy)=\int\mathbb{E}_{P_{2,\theta}}(\mathbb{I}_A|S=y)S_\# P_{1,\theta}(dy)=P_{1,\theta}(A).$$

□

Chapter 4

Asymptotic equivalence of Lévy density estimation and Gaussian white noise

Résumé Au cours du Chapitre 4 nous présentons un résultat d'équivalence asymptotique entre les expériences associées à l'observation discrète (haute fréquence) ou continue d'un processus de Lévy à sauts purs et un modèle de bruit blanc gaussien observé jusqu'à un temps T qui tend vers l'infini. Ici, le paramètre d'intérêt est la densité de Lévy. Toutes les équivalences asymptotiques sont établies en construisant des noyaux de Markov explicites qui peuvent être utilisés pour reproduire une expérience à partir de l'autre. Ce chapitre est basé sur le preprint Mariucci (2015d).

Mot clés: Expériences statistiques non paramétriques, distance de Le Cam, processus de Lévy, densité de Lévy, modèle de bruit blanc gaussien.

Abstract The aim of Chapter 4 is to establish a global asymptotic equivalence between the experiments generated by the discrete (high frequency) or continuous observation of a path of a Lévy process and a Gaussian white noise experiment observed up to a time T, with T tending to infinity. These approximations are given in the sense of the Le Cam distance, under some smoothness conditions on the unknown Lévy density. All the asymptotic equivalences are established by constructing explicit Markov kernels that can be used to reproduce one experiment from the other. This chapter is based on a submitted paper.

Keywords: Nonparametric experiments, Le Cam distance, Lévy processes, Lévy density, Gaussian white noise.

4.1 Introduction

Lévy processes are a fundamental tool in modelling situations, like the dynamics of asset prices and weather measurements, where sudden changes in values may happen. For that reason they are widely employed, among many other fields, in mathematical finance. To name a simple example, the price of a commodity at time t is commonly given as an exponential function of a Lévy process. In general, exponential Lévy models are proposed for their ability to take into account several empirical features observed in the returns of assets such as heavy tails, high-kurtosis and asymmetry (see Cont, Tankov (2004) for an introduction to financial applications).

From a mathematical point of view, Lévy processes are a natural extension of the Brownian motion which preserves the tractable statistical properties of its increments, while relaxing the continuity of paths. The jump dynamics of a Lévy process is dictated by its Lévy density, say f. If f is continuous, its value at a point x_0 determines how frequent jumps of size close to x_0 are to occur per unit time. Concretely, if X is a pure jump Lévy process with Lévy density f, then the function f is such that

$$\int_A f(x)dx = \frac{1}{t}\mathbb{E}\Big[\sum_{s\leq t} \mathbb{I}_A(\Delta X_s)\Big],$$

for any Borel set A and $t > 0$. Here, $\Delta X_s \equiv X_s - X_{s^-}$ denotes the magnitude of the jump of X at time s and $\mathbb{I}_A$ is the characteristic function. Thus, the Lévy measure

$$\nu(A) := \int_A f(x)dx,$$

is the average number of jumps (per unit time) whose magnitudes fall in the set A. Understanding the jumps behavior, therefore requires to estimate the Lévy measure. Several recent works have treated this problem, see e.g. Belomestny et al. (2015) for an overview.

When the available data consists of the whole trajectory of the process during a time interval $[0, T]$, the problem of estimating f may be reduced to estimating the intensity function of an inhomogeneous Poisson process (see, e.g. Figueroa-López, Houdré (2006); Reynaud-Bouret (2003)). However, a continuous-time sampling is never available in practice and thus the relevant problem is that of estimating f based on discrete sample data $X_{t_0}, \dots, X_{t_n}$ during a time interval $[0, T_n]$. In that case, the jumps are latent (unobservable) variables and that clearly adds to the difficulty of the problem. From now on we will place ourselves in a high-frequency setting, that is we assume that the sampling interval $\Delta_n = t_i - t_{i-1}$ tends to zero as n goes to infinity. Such a high-frequency based statistical approach has played a central role in the recent literature on nonparametric estimation for Lévy processes (see e.g. Bec, Lacour (2012); Comte, Genon-Catalot (2010, 2011); Duval (2013); Figueroa-López (2009)). Moreover, in order to make consistent estimation possible, we will also ask the observation time T_n to tend to infinity in order to allow the identification of the jump part in the limit.

Our aim is to prove that, under suitable hypotheses, estimating the Lévy density f is equivalent to estimating the drift of an adequate Gaussian white noise model. In general, asymptotic equivalence results for statistical experiments provide a deeper understanding of statistical problems and allow to single out their main features. The idea is to pass via asymptotic equivalence to another experiment which is easier to analyze. By definition, two sequences of experiments $\mathscr{P}_{1,n}$ and $\mathscr{P}_{2,n}$, defined on possibly different sample spaces, but with the same parameter set, are asymptotically equivalent if the Le Cam distance $\Delta(\mathscr{P}_{1,n}, \mathscr{P}_{2,n})$ tends to zero. For $\mathscr{P}_i = (\mathscr{X}_i, \mathscr{A}_i, \big(P_{i,\theta} : \theta \in \Theta)\big)$, $i = 1, 2$, $\Delta(\mathscr{P}_1, \mathscr{P}_2)$ is the symmetrization of the deficiency $\delta(\mathscr{P}_1, \mathscr{P}_2)$ where

$$\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_K \sup_{\theta \in \Theta} \big\| K P_{1,\theta} - P_{2,\theta} \big\|_{TV}.$$

Here the infimum is taken over all randomizations from $(\mathscr{X}_1, \mathscr{A}_1)$ to $(\mathscr{X}_2, \mathscr{A}_2)$ and $\| \cdot \|_{TV}$ denotes the total variation distance. Roughly speaking, the Le Cam distance quantifies how much one fails to reconstruct (with the help of a randomization) a model from the other one and vice versa. Therefore, we say that $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$ can be interpreted as "the models $\mathscr{P}_1$ and $\mathscr{P}_2$ contain the same amount of information about the parameter θ." The general definition of randomization is quite involved but, in the most frequent examples (namely when the sample spaces are Polish and the experiments dominated), it reduces to that of a Markov kernel. One of the most important features of the Le Cam distance is that it can be also interpreted in terms of statistical decision theory (see Le Cam (1986); Le Cam, Yang (2000); a short review is presented in the Appendix). As a consequence, saying that two statistical models are equivalent means that any statistical inference procedure can be transferred from one model to the other in such a way that the asymptotic risk remains the same, at least for bounded loss functions. Also, as soon as two models, $\mathscr{P}_{1,n}$ and $\mathscr{P}_{2,n}$, that share the same parameter space Θ are proved to be asymptotically equivalent, the same result automatically holds for the restrictions of both $\mathscr{P}_{1,n}$ and $\mathscr{P}_{2,n}$ to a smaller subclass of Θ.

Historically, the first results of asymptotic equivalence in a nonparametric context date from 1996 and are due to Brown, Low (1996) and Nussbaum (1996). The first two authors have shown the asymptotic equivalence of nonparametric regression and a Gaussian white noise model while the third one those of density estimation and white noise. Over the years many generalizations of these results have been proposed such as Brown et al.

(2002b); Carter (2006b, 2007, 2009); Grama, Nussbaum (2002); Meister, Reiß (2013); Reiß (2008); Rohde (2004); Schmidt-Hieber (2014) for nonparametric regression or Brown et al. (2004a); Carter (2002); Jähnisch, Nussbaum (2003) for nonparametric density estimation models. Another very active field of study is that of diffusion experiments. The first result of equivalence between diffusion models and Euler scheme was established in 1998, see Milstein, Nussbaum (1998). In later papers generalizations of this result have been considered (see Genon-Catalot, Laredo (2014); Mariucci (2015c)). Among others we can also cite equivalence results for generalized linear models Grama, Nussbaum (1998), time series Grama, Neumann (2006); Milstein, Nussbaum (1998), diffusion models Dalalyan, Reiß (2006, 2007b); Delattre, Hoffmann (2002); Genon-Catalot, Laredo, Nussbaum (2002), GARCH model Buchmann, Müller (2012), functional linear regression Meister (2011), spectral density estimation Golubev, Nussbaum, Zhou (2010) and volatility estimation Reiß (2011). Negative results are somewhat harder to come by; the most notable among them are Brown, Zhang (1998); Efromovich, Samarov (1996); Wang (2002a). There is however a lack of equivalence results concerning processes with jumps. A first result in this sense is Mariucci (2015b) in which global asymptotic equivalences between the experiments generated by the discrete or continuous observation of a path of a Lévy process and a Gaussian white noise experiment are established. More precisely, in that paper, we have shown that estimating the drift function h from a continuously or discretely (high frequency) time inhomogeneous jump-diffusion process:

$$X_t = \int_0^t h(s)ds + \int_0^t \sigma(s)dW_s + \sum_{i=1}^{N_t} Y_i, \quad t \in [0, T_n], \tag{4.1}$$

is asymptotically equivalent to estimate h in the Gaussian model:

$$dy_t = h(t)dt + \sigma(t)dW_t, \quad t \in [0, T_n].$$

Here we try to push the analysis further and we focus on the case in which the considered parameter is the Lévy density and $X = (X_t)$ is a pure jump Lévy process (see Carr et al. (2002) for the interest of such a class of processes when modelling asset returns). More in detail, we consider the problem of estimating the Lévy density (with respect to a fixed, possibly infinite, Lévy measure ν_0 concentrated on $I \subseteq \mathbb{R}$) $f := \frac{d\nu}{d\nu_0} : I \to \mathbb{R}$ from a continuously or discretely observed pure jump Lévy process X with possibly infinite Lévy measure. Here $I \subseteq \mathbb{R}$ denotes a possibly infinite interval and ν_0 is supposed to be absolutely continuous with respect to Lebesgue with a strictly positive density $g := \frac{d\nu_0}{d\mathrm{Leb}}$. In the case where ν is of finite variation one may write:

$$X_t = \sum_{0<s\leq t} \Delta X_s \tag{4.2}$$

or, equivalently, X has a characteristic function given by:

$$\mathbb{E}\big[e^{iuX_t}\big] = \exp\bigg(-t\bigg(\int_I (1-e^{iuy})\nu(dy)\bigg)\bigg).$$

We suppose that the function f belongs to some a priori set $\mathscr{F}$, nonparametric in general. The discrete observations are of the form X_{t_i}, where $t_i = T_n\frac{i}{n}$, $i = 0,\dots,n$ with $T_n = n\Delta_n \to \infty$ and $\Delta_n \to 0$ as n goes to infinity. We will denote by $\mathscr{P}_n^{\nu_0}$ the statistical model associated with the continuous observation of a trajectory of X until time T_n (which is supposed to go to infinity as n goes to infinity) and by $\mathscr{Q}_n^{\nu_0}$ the one associated with the observation of the discrete data $(X_{t_i})_{i=0}^n$. The aim of this paper is to prove that, under adequate hypotheses on $\mathscr{F}$ (for example, f must be bounded away from zero and infinity; see Section 4.2.1 for a complete definition), the models $\mathscr{P}_n^{\nu_0}$ and $\mathscr{Q}_n^{\nu_0}$ are both asymptotically equivalent to a sequence of Gaussian white noise models of the form:

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n}}\frac{dW_t}{\sqrt{g(t)}}, \quad t \in I.$$

As a corollary, we then get the asymptotic equivalence between $\mathscr{P}_n^{\nu_0}$ and $\mathscr{Q}_n^{\nu_0}$. The main results are precisely stated as Theorems 4.2.5 and 4.2.6. A particular case of special interest arises when X is a compound Poisson process, $\nu_0 \equiv \mathrm{Leb}([0,1])$ and $\mathscr{F} \subseteq \mathscr{F}^I_{(\gamma,K,\kappa,M)}$ where, for fixed $\gamma \in (0,1]$ and K,κ,M strictly positive constants, $\mathscr{F}^I_{(\gamma,K,\kappa,M)}$ is a class of continuously differentiable functions on I defined as follows:

$$\mathscr{F}^I_{(\gamma,K,\kappa,M)} = \Big\{f : \kappa \le f(x) \le M,\ |f'(x)-f'(y)| \le K|x-y|^\gamma,\ \forall x,y \in I\Big\}. \tag{4.3}$$

In this case, the statistical models $\mathscr{P}_n^{\nu_0}$ and $\mathscr{Q}_n^{\nu_0}$ are both equivalent to the Gaussian white noise model:

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n}}dW_t, \quad t \in [0,1].$$

See Example 4.3.1 for more details. By a theorem of Brown and Low in Brown, Low (1996), we obtain, a posteriori, an asymptotic equivalence with the regression model

$$Y_i = \sqrt{f\Big(\frac{i}{T_n}\Big)} + \frac{1}{2\sqrt{T_n}}\xi_i, \quad \xi_i \sim \mathscr{N}(0,1), \quad i = 1,\dots,[T_n].$$

Note that a similar form of a Gaussian shift was found to be asymptotically equivalent to a nonparametric density estimation experiment, see Nussbaum (1996). Let us mention that we also treat some explicit examples where ν_0 is neither finite nor compactly-supported (see Examples 4.3.2 and 4.3.3).

Without entering into any detail, we remark here that the methods are very different from those in Mariucci (2015b). In particular, since f belongs to the discontinuous part of a Lévy process, rather than its continuous part, the Girsanov-type changes of measure are irrelevant here. We thus need new instruments, like the Esscher changes of measure.

Our proof is based on the construction, for any given Lévy measure ν, of two adequate approximations $\hat{\nu}_m$ and $\bar{\nu}_m$ of ν: the idea of discretizing the Lévy density already appeared in an earlier work, see Etoré, Louhichi, Mariucci (2013). The present work is also inspired by the papers Carter (2002) (for a multinomial approximation), Brown et al. (2004a) (for passing from independent Poisson variables to independent normal random variables) and Mariucci (2015b) (for a Bernoulli approximation). This method allows us to construct explicit Markov kernels that lead from one model to the other; these may be applied in practice to transfer minimax estimators.

The paper is organized as follows: Sections 4.2.1 and 4.2.2 are devoted to make the parameter space and the considered statistical experiments precise. The main results are given in Section 4.2.3, followed by Section 4.3 in which some examples can be found. The proofs are postponed to Section 4.4. The paper includes an Appendix recalling the definition and some useful properties of the Le Cam distance as well as of Lévy processes.

4.2 Assumptions and main results

4.2.1 The parameter space

Consider a (possibly infinite) Lévy measure ν_0 concentrated on a possibly infinite interval $I \subseteq \mathbb{R}$, admitting a density $g > 0$ with respect to Lebesgue. The parameter space of the experiments we are concerned with is a class of functions $\mathscr{F} = \mathscr{F}^{\nu_0, I}$ defined on I that form a class of Lévy densities with respect to ν_0: For each $f \in \mathscr{F}$, let ν (resp. $\hat{\nu}_m$) be the Lévy measure having f (resp. $\hat{f}_m$) as a density with respect to ν_0 where, for every $f \in \mathscr{F}$, $\hat{f}_m(x)$ is defined as follows.

Suppose first $x > 0$. Given a positive integer depending on n, $m = m_n$, let $J_j := (v_{j-1}, v_j]$ where $v_1 = \varepsilon_m \geq 0$ and v_j are chosen in such a way that

$$\mu_m := \nu_0(J_j) = \frac{\nu_0\big((I \setminus [0, \varepsilon_m]) \cap \mathbb{R}_+\big)}{m-1}, \quad \forall j = 2, \dots, m. \tag{4.4}$$

In the sequel, for the sake of brevity, we will only write m without making explicit the dependence on n. Define $x_j^* := \frac{\int_{J_j} x \nu_0(dx)}{\mu_m}$ and introduce a sequence of functions $0 \leq V_j \leq \frac{1}{\mu_m}$, $j = 2, \dots, m$ supported on $[x_{j-1}^*, x_{j+1}^*]$ if $j = 3, \dots, m-1$, on $[\varepsilon_m, x_3^*]$ if

$j = 2$ and on $(I \setminus [0, x^*_{m-1}]) \cap \mathbb{R}_+$ if $j = m$. The V_j's are defined recursively in the following way.

- V_2 is equal to $\frac{1}{\mu_m}$ on the interval $(\varepsilon_m, x^*_2]$ and on the interval $(x^*_2, x^*_3]$ it is chosen so that it is continuous (in particular, $V_2(x^*_2) = \frac{1}{\mu_m}$), $\int_{x^*_2}^{x^*_3} V_2(y)\nu_0(dy) = \frac{\nu_0((x^*_2, v_2]))}{\mu_m}$ and $V_2(x^*_3) = 0$.

- For $j = 3, \dots, m-1$ define V_j as the function $\frac{1}{\mu_m} - V_{j-1}$ on the interval $[x^*_{j-1}, x^*_j]$. On $[x^*_j, x^*_{j+1}]$ choose V_j continuous and such that $\int_{x^*_j}^{x^*_{j+1}} V_j(y)\nu_0(dy) = \frac{\nu_0((x^*_j, v_j]))}{\mu_m}$ and $V_j(x^*_{j+1}) = 0$.

- Finally, let V_m be the function supported on $(I \setminus [0, x^*_{m-1}]) \cap \mathbb{R}_+$ such that

$$V_m(x) = \frac{1}{\mu_m} - V_{m-1}(x), \quad \text{for } x \in [x^*_{m-1}, x^*_m],$$
$$V_m(x) = \frac{1}{\mu_m}, \quad \text{for } x \in (I \setminus [0, x^*_m]) \cap \mathbb{R}_+.$$

(It is immediate to check that such a choice is always possible). Observe that, by construction,

$$\sum_{j=2}^{m} V_j(x)\mu_m = 1, \quad \forall x \in (I \setminus [0, \varepsilon_m]) \cap \mathbb{R}_+ \quad \text{and} \quad \int_{(I \setminus [0,\varepsilon_m]) \cap \mathbb{R}_+} V_j(y)\nu_0(dy) = 1.$$

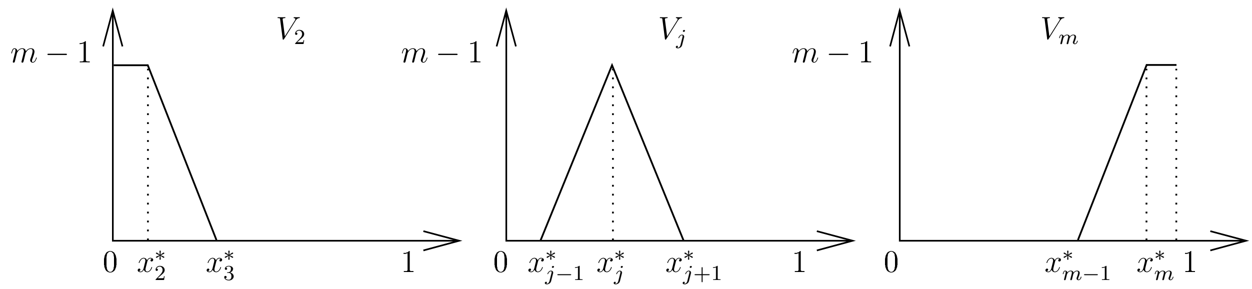
Analogously, define $\mu^-_m = \frac{\nu_0\left((I \setminus [-\varepsilon_m, 0]) \cap \mathbb{R}_-\right)}{m-1}$ and $J_{-m}, \dots, J_{-2}$ such that $\nu_0(J_{-j}) = \mu^-_m$ for all j. Then, for $x < 0$, x^*_{-j} is defined as x^*_j by using J_{-j} and μ^-_m instead of J_j and μ_m and the V_{-j}'s are defined with the same procedure as the V_j's, starting from V_{-2} and proceeding by induction.

Define

$$\hat{f}_m(x) = \mathbb{I}_{[-\varepsilon_m, \varepsilon_m]}(x) + \sum_{j=2}^{m} \left(V_j(x) \int_{J_j} f(y)\nu_0(dy) + V_{-j}(x) \int_{J_{-j}} f(y)\nu_0(dy) \right). \tag{4.5}$$

The definitions of the V_j's above are modeled on the following example:

Example 4.2.1. Let ν_0 be the Lebesgue measure on $[0, 1]$ and $\varepsilon_m = 0$. Then $v_j = \frac{j-1}{m-1}$ and $x^*_j = \frac{2j-3}{2m-2}$, $j = 2, \dots, m$. The standard choice for V_j (based on the construction by Carter (2002)) is given by the piecewise linear functions interpolating the values in the points x^*_j specified above:



Remark 4.2.2. The function $\hat{f}_m$ has been defined in such a way that the rate of convergence of the L_2 norm between the restriction of f and $\hat{f}_m$ on $I\setminus[-\varepsilon_m,\varepsilon_m]$ is compatible with the rate of convergence of the other quantities appearing in the statements of Theorems 4.2.5 and 4.2.6. For that reason, as in Carter (2002), we have not chosen a piecewise constant approximation of f but an approximation that is, at least in the simplest cases, a piecewise linear approximation of f. Such a choice allows us to gain an order of magnitude on the convergence rate of $\|f-\hat{f}_m\|_{L_2(\nu_0|I\setminus[-\varepsilon_m,\varepsilon_m])}$ at least when $\mathscr{F}$ is a class of sufficiently smooth functions.

We now explain the assumptions we will need to make on the parameter $f\in\mathscr{F}=\mathscr{F}^{\nu_0,I}$. The superscripts ν_0 and I will be suppressed whenever this can lead to no confusion. We require that:

(H1) There exist constants $\kappa, M>0$ such that $\kappa\leq f(y)\leq M$, for all $y\in I$ and $f\in\mathscr{F}$.

For every integer $m=m_n$, we can consider $\widehat{\sqrt{f}}_m$, the approximation of $\sqrt{f}$ constructed as $\hat{f}_m$ above, i.e. $\widehat{\sqrt{f}}_m(x)=\mathbb{I}_{[-\varepsilon_m,\varepsilon_m]}(x)+\sum_{\substack{j=-m\dots,m\\ j\neq-1,0,1.}}V_j(x)\int_{J_j}\sqrt{f(y)}\nu_0(dy)$, and introduce the quantities:

$$A_m^2(f):=\int_{I\setminus\big[-\varepsilon_m,\varepsilon_m\big]}\Big(\widehat{\sqrt{f}}_m(y)-\sqrt{f(y)}\Big)^2\nu_0(dy),$$

$$B_m^2(f):=\sum_{\substack{j=-m\dots,m\\ j\neq-1,0,1.}}\bigg(\int_{J_j}\frac{\sqrt{f(y)}}{\sqrt{\nu_0(J_j)}}\nu_0(dy)-\sqrt{\nu(J_j)}\bigg)^2,$$

$$C_m^2(f):=\int_{-\varepsilon_m}^{\varepsilon_m}\big(\sqrt{f(t)}-1\big)^2\nu_0(dt).$$

The conditions defining the parameter space $\mathscr{F}$ are expressed by asking that the quantities introduced above converge quickly enough to zero. To state the assumptions of Theorem 4.2.5 precisely, we will assume the existence of sequences of discretizations $m=m_n\to\infty$, of positive numbers $\varepsilon_m=\varepsilon_{m_n}\to 0$ and of functions V_j, $j=\pm 2,\dots,\pm m$, such that:

(C1) $\lim_{n\to\infty} n\Delta_n \sup_{f\in\mathscr{F}} \int_{I\setminus(-\varepsilon_m,\varepsilon_m)} \Big(f(x)-\hat{f}_m(x)\Big)^2 \nu_0(dx) = 0.$

(C2) $\lim_{n\to\infty} n\Delta_n \sup_{f\in\mathscr{F}} \big(A_m^2(f)+B_m^2(f)+C_m^2(f)\big) = 0.$

Remark in particular that Condition (C2) implies the following:

(H2) $\sup_{f\in\mathscr{F}} \int_I (\sqrt{f(y)}-1)^2 \nu_0(dy) \leq L,$

where $L = \sup_{f\in\mathscr{F}} \int_{-\varepsilon_m}^{\varepsilon_m} (\sqrt{f(x)}-1)^2\nu_0(dx) + (\sqrt{M}+1)^2\nu_0\big(I\setminus(-\varepsilon_m,\varepsilon_m)\big)$, for any choice of m such that the quantity in the limit appearing in Condition (C2) is finite.

Theorem 4.2.6 has slightly stronger hypotheses, defining possibly smaller parameter spaces: We will assume the existence of sequences m_n, ε_m and V_j, $j=\pm 2,\dots,\pm m$ (possibly different from the ones above) such that Condition (C1) is verified and the following stronger version of Condition (C2) holds:

(C2') $\lim_{n\to\infty} n\Delta_n \sup_{f\in\mathscr{F}} \big(A_m^2(f)+B_m^2(f)+nC_m^2(f)\big) = 0.$

Finally, some of our results have a more explicit statement under the hypothesis of finite variation which we state as:

(FV) $\int_I (|x|\wedge 1)\nu_0(dx) < \infty.$

Remark 4.2.3. The Condition (C1) and those involving the quantities $A_m(f)$ and $B_m(f)$ all concern similar but slightly different approximations of f. In concrete examples, they may all be expected to have the same rate of convergence but to keep the greatest generality we preferred to state them separately. On the other hand, conditions on the quantity $C_m(f)$ are purely local around zero, requiring, for each $f\in\mathscr{F}$, that $f(x)$ tends to 1 quickly enough as x tends to 0.

Examples 4.2.4. To get a grasp on Conditions (C1), (C2) we analyze here three different examples according to the different behavior of ν_0 near $0\in I$. In all of these cases the parameter space $\mathscr{F}^{\nu_0,I}$ will be a subclass of $\mathscr{F}^I_{(\gamma,K,\kappa,M)}$ defined as in (4.3). Recall that the conditions (C1), (C2) and (C2') depend on the choice of sequences m_n, ε_m and functions V_j. For the first two of the three examples, where $I=[0,1]$, we will make the standard choice for V_j of triangular and trapezoidal functions, similarly to those in Example 4.2.1. Namely, for $j=3,\dots,m-1$ we have

$$V_j(x) = \mathbb{I}_{(x_{j-1}^*,x_j^*]}(x)\frac{x-x_{j-1}^*}{x_j^*-x_{j-1}^*}\frac{1}{\mu_m} + \mathbb{I}_{(x_j^*,x_{j+1}^*]}(x)\frac{x_{j+1}^*-x}{x_{j+1}^*-x_j^*}\frac{1}{\mu_m}; \tag{4.6}$$

the two extremal functions V_2 and V_m are chosen so that $V_2 \equiv \frac{1}{\mu_m}$ on $(\varepsilon_m, x_2^*]$ and $V_m \equiv \frac{1}{\mu_m}$ on $(x_m^*, 1]$. In the second example, where ν_0 is infinite, one is forced to take $\varepsilon_m > 0$ and to keep in mind that the x_j^* are not uniformly distributed on $[\varepsilon_m, 1]$. Proofs of all the statements here can be found in Section 4.5.2.

1. The finite case: $\nu_0 \equiv \mathrm{Leb}([0,1])$.

In this case we are free to choose $\mathscr{F}^{\mathrm{Leb},[0,1]} = \mathscr{F}_{(\gamma,K,\kappa,M)}^{[0,1]}$. Indeed, as ν_0 is finite, there is no need to single out the first interval $J_1 = [0, \varepsilon_m]$, so that $C_m(f)$ does not enter in the proofs and the definitions of $A_m(f)$ and $B_m(f)$ involve integrals on the whole of $[0,1]$. Also, the choice of the V_j's as in (4.6) guarantees that $\int_0^1 V_j(x)dx = 1$. Then, the quantities $\|f - \hat{f}_m\|_{L_2([0,1])}$, $A_m(f)$ and $B_m(f)$ all have the same rate of convergence, which is given by:

$$\sqrt{\int_0^1 \Big(f(x) - \hat{f}_m(x)\Big)^2 \nu_0(dx)} + A_m(f) + B_m(f) = O\Big(m^{-\gamma-1} + m^{-\frac{3}{2}}\Big),$$

uniformly on f. See Section 4.5.2 for a proof.

2. The finite variation case: $\frac{d\nu_0}{d\mathrm{Leb}}(x) = x^{-1}\mathbb{I}_{[0,1]}(x)$.

In this case, the parameter space $\mathscr{F}^{\nu_0,[0,1]}$ is a proper subset of $\mathscr{F}_{(\gamma,K,\kappa,M)}^{[0,1]}$. Indeed, as we are obliged to choose $\varepsilon_m > 0$, we also need to impose that $C_m(f) = o\big(\frac{1}{n\sqrt{\Delta_n}}\big)$, with uniform constants with respect to f, that is, that all $f \in \mathscr{F}$ converge to 1 quickly enough as $x \to 0$. Choosing $\varepsilon_m = m^{-1-\alpha}$, $\alpha > 0$ we have that $\mu_m = \frac{\ln(\varepsilon_m^{-1})}{m-1}$, $v_j = \varepsilon_m^{\frac{m-j}{m-1}}$ and $x_j^* = \frac{(v_j - v_{j-1})}{\mu_m}$. In particular, $\max_j |v_{j-1} - v_j| = |v_m - v_{m-1}| = O\Big(\frac{\ln m}{m}\Big)$. Also in this case one can prove that the standard choice of V_j described above leads to $\int_{\varepsilon_m}^1 V_j(x)\frac{dx}{x} = 1$. Again, the quantities $\|f - \hat{f}_m\|_{L_2(\nu_0|I\setminus[0,\varepsilon_m])}$, $A_m(f)$ and $B_m(f)$ have the same rate of convergence given by:

$$\sqrt{\int_{\varepsilon_m}^1 \Big(f(x) - \hat{f}_m(x)\Big)^2 \nu_0(dx)} + A_m(f) + B_m(f) = O\bigg(\Big(\frac{\ln m}{m}\Big)^{\gamma+1} \sqrt{\ln(\varepsilon_m^{-1})}\bigg), \tag{4.7}$$

uniformly on f. The condition on $C_m(f)$ depends on the behavior of f near 0. For example, it is ensured if one considers a parametric family of the form $f(x) = e^{-\lambda x}$ with a bounded $\lambda > 0$. See Section 4.5.2 for a proof.

3. The infinite variation, non-compactly supported case: $\frac{d\nu_0}{d\mathrm{Leb}}(x) = x^{-2}\mathbb{I}_{\mathbb{R}_+}(x)$.

This example involves significantly more computations than the preceding ones, since the classical triangular choice for the functions V_j would not have integral equal to 1 (with respect to ν_0), and the support is not compact. The parameter space $\mathscr{F}^{\nu_0,[0,\infty)}$ can still be chosen as a proper subclass of $\mathscr{F}_{(\gamma,K,\kappa,M)}^{[0,\infty)}$, again by imposing that $C_m(f)$ converges to zero quickly enough (more details about this condition are discussed in Example 4.3.3). We divide the interval $[0,\infty)$ in m intervals $J_j=[v_{j-1},v_j)$ with:

$$v_0=0;\quad v_1=\varepsilon_m;\quad v_j=\frac{\varepsilon_m(m-1)}{m-j};\quad v_m=\infty;\quad \mu_m=\frac{1}{\varepsilon_m(m-1)}.$$

To deal with the non-compactness problem, we choose some "horizon" $H(m)$ that goes to infinity slowly enough as m goes to infinity and we bound the L_2 distance between f and $\hat{f}_m$ for $x>H(m)$ by $2\sup_{x\geq H(m)}\frac{f(x)^2}{H(m)}$. We have:

$$\|f-\hat{f}_m\|^2_{L_2(\nu_0|I\setminus[0,\varepsilon_m])}+A_m^2(f)+B_m^2(f)=O\bigg(\frac{H(m)^{3+4\gamma}}{(\varepsilon_m m)^{2+2\gamma}}+\sup_{x\geq H(m)}\frac{f(x)^2}{H(m)}\bigg).$$

In the general case where the best estimate for $\sup_{x\geq H(m)} f(x)^2$ is simply given by M^2, an optimal choice for $H(m)$ is $\sqrt{\varepsilon_m m}$, that gives a rate of convergence:

$$\|f-\hat{f}_m\|^2_{L_2(\nu_0|I\setminus[0,\varepsilon_m])}+A_m^2(f)+B_m^2(f)=O\bigg(\frac{1}{\sqrt{\varepsilon_m m}}\bigg),$$

independently of γ. See Section 4.5.2 for a proof.

4.2.2 Definition of the experiments

Let $(x_t)_{t\geq 0}$ be the canonical process on the Skorokhod space $(D,\mathscr{D})$ and denote by $P^{(b,0,\nu)}$ the law induced on $(D,\mathscr{D})$ by a Lévy process with characteristic triplet $(b,0,\nu)$. We will write $P_t^{(b,0,\nu)}$ for the restriction of $P^{(b,0,\nu)}$ to the σ-algebra $\mathscr{D}_t$ generated by $\{x_s:0\leq s\leq t\}$ (see 4.5.5 for the precise definitions). Let $Q_t^{(b,0,\nu)}$ be the marginal law at time t of a Lévy process with characteristic triplet $(b,0,\nu)$. In the case where $\int_{|y|\leq 1}|y|\nu(dy)<\infty$ we introduce the notation $\gamma^\nu:=\int_{|y|\leq 1}y\nu(dy)$; then, Condition (H2) guarantees the finiteness of $\gamma^{\nu-\nu_0}$ (see Remark 33.3 in Sato (1999) for more details).

Recall that we introduced the discretization $t_i=T_n\frac{i}{n}$ of $[0,T_n]$ and denote by $\mathbf{Q}_n^{(\gamma^{\nu-\nu_0},0,\nu)}$ the laws of the $n+1$ marginals of $(x_t)_{t\geq 0}$ at times t_i, $i=0,\dots,n$. We will consider the following statistical models, depending on a fixed, possibly infinite, Lévy measure ν_0 concentrated on I (clearly, the models with the subscript FV are meaningful only under the assumption (FV)):

$$\begin{aligned}
\mathscr{P}_{n,FV}^{\nu_0} &= \bigg(D, \mathscr{D}_{T_n}, \Big\{P_{T_n}^{(\gamma^\nu,0,\nu)} : f := \frac{d\nu}{d\nu_0} \in \mathscr{F}^{\nu_0,I}\Big\}\bigg),\\
\mathscr{Q}_{n,FV}^{\nu_0} &= \bigg(\mathbb{R}^{n+1}, \mathscr{B}(\mathbb{R}^{n+1}), \Big\{\mathbf{Q}_n^{(\gamma^\nu,0,\nu)} : f := \frac{d\nu}{d\nu_0} \in \mathscr{F}^{\nu_0,I}\Big\}\bigg),\\
\mathscr{P}_{n}^{\nu_0} &= \bigg(D, \mathscr{D}_{T_n}, \Big\{P_{T_n}^{(\gamma^{\nu-\nu_0},0,\nu)} : f := \frac{d\nu}{d\nu_0} \in \mathscr{F}^{\nu_0,I}\Big\}\bigg),\\
\mathscr{Q}_{n}^{\nu_0} &= \bigg(\mathbb{R}^{n+1}, \mathscr{B}(\mathbb{R}^{n+1}), \Big\{\mathbf{Q}_n^{(\gamma^{\nu-\nu_0},0,\nu)} : f := \frac{d\nu}{d\nu_0} \in \mathscr{F}^{\nu_0,I}\Big\}\bigg).
\end{aligned}$$

Finally, let us introduce the Gaussian white noise model that will appear in the statement of our main results. For that, let us denote by $(C(I),\mathscr{C})$ the space of continuous mappings from I into $\mathbb{R}$ endowed with its standard filtration, by g the density of ν_0 with respect to the Lebesgue measure. We will require $g>0$ and let $\mathbb{W}_n^f$ be the law induced on $(C(I),\mathscr{C})$ by the stochastic process satisfying:

$$dy_t = \sqrt{f(t)}dt + \frac{dW_t}{2\sqrt{T_n}\sqrt{g(t)}}, \quad t \in I, \tag{4.8}$$

where $(W_t)_{t\in\mathbb{R}}$ denotes a Brownian motion on $\mathbb{R}$ with $W_0=0$. Then we set:

$$\mathscr{W}_n^{\nu_0} = \Big(C(I), \mathscr{C}, \{\mathbb{W}_n^f : f \in \mathscr{F}^{\nu_0,I}\}\Big).$$

Observe that when ν_0 is a finite Lévy measure, then $\mathscr{W}_n^{\nu_0}$ is equivalent to the statistical model associated with the continuous observation of a process $(\tilde{y}_t)_{t\in I}$ defined by:

$$d\tilde{y}_t = \sqrt{f(t)g(t)}dt + \frac{dW_t}{2\sqrt{T_n}}, \quad t \in I.$$

4.2.3 Main results

Using the notation introduced in Section 4.2.1, we now state our main results. For brevity of notation, we will denote by $H(f,\hat{f}_m)$ (resp. $L_2(f,\hat{f}_m)$) the Hellinger distance (resp. the L_2 distance) between the Lévy measures ν and $\hat{\nu}_m$ restricted to $I \setminus [-\varepsilon_m, \varepsilon_m]$, i.e.:

$$\begin{aligned}
H^2(f,\hat{f}_m) &:= \int_{I\setminus[-\varepsilon_m,\varepsilon_m]} \Big(\sqrt{f(x)} - \sqrt{\hat{f}_m(x)}\Big)^2 \nu_0(dx),\\
L_2(f,\hat{f}_m)^2 &:= \int_{I\setminus[-\varepsilon_m,\varepsilon_m]} \big(f(y) - \hat{f}_m(y)\big)^2 \nu_0(dy).
\end{aligned}$$

Observe that Condition (H1) implies (see Lemma 4.5.1)

$$\frac{1}{4M} L_2(f,\hat{f}_m)^2 \le H^2(f,\hat{f}_m) \le \frac{1}{4\kappa} L_2(f,\hat{f}_m)^2.$$

Theorem 4.2.5. *Let ν_0 be a known Lévy measure concentrated on a (possibly infinite) interval $I \subseteq \mathbb{R}$ and having strictly positive density with respect to the Lebesgue measure. Let us choose a parameter space $\mathscr{F}^{\nu_0,I}$ such that there exist a sequence $m = m_n$ of integers, functions V_j, $j = \pm 2, \dots, \pm m$ and a sequence $\varepsilon_m \to 0$ as $m \to \infty$ such that Conditions* (H1), (C1), (C2) *are satisfied for $\mathscr{F} = \mathscr{F}^{\nu_0,I}$. Then, for n big enough we have:*

$$\Delta(\mathscr{P}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\bigg(\sqrt{n\Delta_n} \sup_{f\in\mathscr{F}} \Big(A_m(f) + B_m(f) + C_m(f)\Big)\bigg) + O\bigg(\sqrt{n\Delta_n} \sup_{f\in\mathscr{F}} L_2(f, \hat{f}_m) + \sqrt{\frac{m}{n\Delta_n}\Big(\frac{1}{\mu_m} + \frac{1}{\mu_m^-}\Big)}\bigg). \tag{4.9}$$

Theorem 4.2.6. *Let ν_0 be a known Lévy measure concentrated on a (possibly infinite) interval $I \subseteq \mathbb{R}$ and having strictly positive density with respect to the Lebesgue measure. Let us choose a parameter space $\mathscr{F}^{\nu_0,I}$ such that there exist a sequence $m = m_n$ of integers, functions V_j, $j = \pm 2, \dots, \pm m$ and a sequence $\varepsilon_m \to 0$ as $m \to \infty$ such that Conditions* (H1), (C1), (C2') *are satisfied for $\mathscr{F} = \mathscr{F}^{\nu_0,I}$. Then, for n big enough we have:*

$$\Delta(\mathscr{Q}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\bigg(\nu_0\Big(I \setminus [-\varepsilon_m, \varepsilon_m]\Big)\sqrt{n\Delta_n^2} + \frac{m\ln m}{\sqrt{n}} + \sqrt{n\sqrt{\Delta_n}\sup_{f\in\mathscr{F}} C_m(f)}\bigg) + O\bigg(\sqrt{n\Delta_n}\sup_{f\in\mathscr{F}}\Big(A_m(f) + B_m(f) + H(f, \hat{f}_m)\Big)\bigg). \tag{4.10}$$

Corollary 4.2.7. *Let ν_0 be as above and let us choose a parameter space $\mathscr{F}^{\nu_0,I}$ so that there exist sequences m_n', ε_m', V_j' and m_n'', ε_m'', V_j'' such that:*

- *Conditions (H1), (C1) and (C2) hold for m_n', ε_m', V_j', and $\frac{m'}{n\Delta_n}\Big(\frac{1}{\mu_{m'}} + \frac{1}{\mu_{m'}^-}\Big)$ tends to zero.*
- *Conditions (H1), (C1) and (C2') hold for m_n'', ε_m'', V_j'', and $\nu_0\Big(I\setminus[-\varepsilon_{m''}, \varepsilon_{m''}]\Big)\sqrt{n\Delta_n^2} + \frac{m''\ln m''}{\sqrt{n}}$ tends to zero.*

Then the statistical models $\mathscr{P}_n^{\nu_0}$ and $\mathscr{Q}_n^{\nu_0}$ are asymptotically equivalent:

$$\lim_{n\to\infty} \Delta(\mathscr{P}_n^{\nu_0}, \mathscr{Q}_n^{\nu_0}) = 0,$$

If, in addition, the Lévy measures have finite variation, i.e. if we assume (FV), then the same results hold replacing $\mathscr{P}_n^{\nu_0}$ and $\mathscr{Q}_n^{\nu_0}$ by $\mathscr{P}_{n,FV}^{\nu_0}$ and $\mathscr{Q}_{n,FV}^{\nu_0}$, respectively (see Lemma 4.5.19).

4.3 Examples

We will now analyze three different examples, underlining the different behaviors of the Lévy measure ν_0 (respectively, finite, infinite with finite variation and infinite with infinite variation). The three chosen Lévy measures are $\mathbb{I}_{[0,1]}(x)dx$, $\mathbb{I}_{[0,1]}(x)\frac{dx}{x}$ and $\mathbb{I}_{\mathbb{R}_+}(x)\frac{dx}{x^2}$. In all three cases we assume the parameter f to be uniformly bounded and with uniformly γ-Hölder derivatives: We will describe adequate subclasses $\mathscr{F}^{\nu_0,I} \subseteq \mathscr{F}^I_{(\gamma,K,\kappa,M)}$ defined as in (4.3). It seems very likely that the same results that are highlighted in these examples hold true for more general Lévy measures; however, we limit ourselves to these examples in order to be able to explicitly compute the quantities involved (v_j, x_j^*, etc.) and hence estimate the distance between f and $\hat{f}_m$ as in Examples 4.2.4.

In the first of the three examples, where ν_0 is the Lebesgue measure on $I = [0,1]$, we are considering the statistical models associated with the discrete and continuous observation of a compound Poisson process with Lévy density f. Observe that $\mathscr{W}_n^{\mathrm{Leb}}$ reduces to the statistical model associated with the continuous observation of a trajectory from:

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n}}dW_t, \quad t \in [0,1].$$

In this case we have:

Example 4.3.1. (Finite Lévy measure). Let ν_0 be the Lebesgue measure on $I = [0,1]$ and let $\mathscr{F} = \mathscr{F}^{\mathrm{Leb},[0,1]}$ be any subclass of $\mathscr{F}^{[0,1]}_{(\gamma,K,\kappa,M)}$ for some strictly positive constants K, κ, M and $\gamma \in (0,1]$. Then:

$$\lim_{n\to\infty} \Delta(\mathscr{P}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}_n^{\mathrm{Leb}}) = 0 \quad \text{and} \quad \lim_{n\to\infty} \Delta(\mathscr{Q}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}_n^{\mathrm{Leb}}) = 0.$$

More precisely,

$$\Delta(\mathscr{P}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}_n^{\mathrm{Leb}}) = \begin{cases} O\Big((n\Delta_n)^{-\frac{\gamma}{4+2\gamma}}\Big) & \text{if } \gamma \in \big(0,\frac{1}{2}\big], \\ O\Big((n\Delta_n)^{-\frac{1}{10}}\Big) & \text{if } \gamma \in \big(\frac{1}{2},1\big]. \end{cases}$$

In the case where $\Delta_n = n^{-\beta}$, $\frac{1}{2} < \beta < 1$, an upper bound for the rate of convergence of $\Delta(\mathscr{Q}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}_n^{\mathrm{Leb}})$ is

$$\Delta(\mathscr{Q}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}_n^{\mathrm{Leb}}) = \begin{cases} O\Big(n^{-\frac{\gamma+\beta}{4+2\gamma}}\ln n\Big) & \text{if } \gamma \in \big(0,\frac{1}{2}\big) \text{ and } \frac{2+2\gamma}{3+2\gamma} \le \beta < 1, \\ O\Big(n^{\frac{1}{2}-\beta}\ln n\Big) & \text{if } \gamma \in \big(0,\frac{1}{2}\big) \text{ and } \frac{1}{2} < \beta < \frac{2+2\gamma}{3+2\gamma}, \\ O\Big(n^{-\frac{2\beta+1}{10}}\ln n\Big) & \text{if } \gamma \in \big[\frac{1}{2},1\big] \text{ and } \frac{3}{4} \le \beta < 1, \\ O\Big(n^{\frac{1}{2}-\beta}\ln n\Big) & \text{if } \gamma \in \big[\frac{1}{2},1\big] \text{ and } \frac{1}{2} < \beta < \frac{3}{4}. \end{cases}$$

See Section 4.5.3 for a proof.

Example 4.3.2. (Infinite Lévy measure with finite variation). Let X be a truncated Gamma process with (infinite) Lévy measure of the form:

$$\nu(A) = \int_A \frac{e^{-\lambda x}}{x} dx, \quad A \in \mathscr{B}([0,1]).$$

Here $\mathscr{F}^{\nu_0, I}$ is a 1-dimensional parametric family in λ, assuming that there exists a known constant λ_0 such that $0 < \lambda \leq \lambda_0 < \infty$, $f(t) = e^{-\lambda t}$ and $d\nu_0(x) = \frac{1}{x}dx$. In particular, f is Lipschitz, i.e. $\mathscr{F}^{\nu_0,[0,1]} \subset \mathscr{F}^{[0,1]}_{(\gamma=1,K,\kappa,M)}$. The discrete or continuous observations (up to time T_n) of X are asymptotically equivalent to $\mathscr{W}_n^{\nu_0}$, the statistical model associated with the observation of a trajectory of the process (y_t):

$$dy_t = \sqrt{f(t)}dt + \frac{\sqrt{t}dW_t}{2\sqrt{T_n}}, \quad t \in [0,1].$$

More precisely, in the case where $\Delta_n = n^{-\beta}$, $\frac{1}{2} < \beta < 1$, an upper bound for the rate of convergence of $\Delta(\mathscr{Q}_{n,FV}^{\nu_0}, \mathscr{W}_n^{\nu_0})$ is

$$\Delta(\mathscr{Q}_{n,FV}^{\nu_0}, \mathscr{W}_n^{\nu_0}) = \begin{cases} O\big(n^{\frac{1}{2}-\beta} \ln n\big) & \text{if } \frac{1}{2} < \beta \leq \frac{9}{10} \\ O\big(n^{-\frac{1+2\beta}{7}} \ln n\big) & \text{if } \frac{9}{10} < \beta < 1. \end{cases}$$

Concerning the continuous setting we have:

$$\Delta(\mathscr{P}_{n,FV}^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\Big(n^{\frac{\beta-1}{6}} \big(\ln n\big)^{\frac{5}{2}}\Big) = O\Big(T_n^{-\frac{1}{6}} \big(\ln T_n\big)^{\frac{5}{2}}\Big).$$

See Section 4.5.4 for a proof.

Example 4.3.3. (Infinite Lévy measure, infinite variation). Let X be a pure jump Lévy process with infinite Lévy measure of the form:

$$\nu(A) = \int_A \frac{2 - e^{-\lambda x^3}}{x^2} dx, \quad A \in \mathscr{B}(\mathbb{R}^+).$$

Again, we are considering a parametric family in $\lambda > 0$, assuming that the parameter stays bounded below a known constant λ_0. Here, $f(t) = 2 - e^{-\lambda t^3}$, hence $1 \leq f(t) \leq 2$, for all $t \geq 0$, and f is Lipschitz, i.e. $\mathscr{F}^{\nu_0,\mathbb{R}_+} \subset \mathscr{F}^{\mathbb{R}_+}_{(\gamma=1,K,\kappa,M)}$. The discrete or continuous observations (up to time T_n) of X are asymptotically equivalent to the statistical model associated with the observation of a trajectory of the process (y_t):

$$dy_t = \sqrt{f(t)}dt + \frac{tdW_t}{2\sqrt{T_n}}, \quad t \geq 0.$$

More precisely, in the case where $\Delta_n = n^{-\beta}$, $0<\beta<1$, an upper bound for the rate of convergence of $\Delta(\mathscr{Q}_n^{\nu_0}, \mathscr{W}_n^{\nu_0})$ is

$$\Delta(\mathscr{Q}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) = \begin{cases} O\big(n^{\frac{1}{2}-\frac{2}{3}\beta}\big) & \text{if } \frac{3}{4}<\beta<\frac{12}{13} \\ O\big(n^{-\frac{1}{6}+\frac{\beta}{18}}(\ln n)^{\frac{7}{6}}\big) & \text{if } \frac{12}{13}\le\beta<1. \end{cases}$$

In the continuous setting, we have

$$\Delta(\mathscr{P}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\big(n^{\frac{3\beta-3}{34}}(\ln n)^{\frac{7}{6}}\big) = O\big(T_n^{-\frac{3}{34}}(\ln T_n)^{\frac{7}{6}}\big).$$

See Section 4.5.5 for a proof.

4.4 Proofs of the main results

In order to simplify notations, the proofs will be presented in the case $I \subseteq \mathbb{R}^+$. Nevertheless, this allows us to present all the main difficulties, since they can only appear near 0. To prove Theorems 4.2.5 and 4.2.6 we need to introduce several intermediate statistical models. In that regard, let us denote by Q_j^f the law of a Poisson random variable with mean $T_n\nu(J_j)$ (see (4.4) for the definition of J_j). We will denote by $\mathscr{L}_m$ the statistical model associated with the family of probabilities $\big\{\bigotimes_{j=2}^m Q_j^f : f\in\mathscr{F}\big\}$:

$$\mathscr{L}_m = \bigg(\bar{\mathbb{N}}^{m-1}, \mathcal{P}(\bar{\mathbb{N}}^{m-1}), \bigg\{\bigotimes_{j=2}^m Q_j^f : f\in\mathscr{F}\bigg\}\bigg). \tag{4.11}$$

By N_j^f we mean the law of a Gaussian random variable $\mathscr{N}(2\sqrt{T_n\nu(J_j)},1)$ and by $\mathscr{N}_m$ the statistical model associated with the family of probabilities $\big\{\bigotimes_{j=2}^m N_j^f : f\in\mathscr{F}\big\}$:

$$\mathscr{N}_m = \bigg(\mathbb{R}^{m-1}, \mathscr{B}(\mathbb{R}^{m-1}), \bigg\{\bigotimes_{j=2}^m N_j^f : f\in\mathscr{F}\bigg\}\bigg). \tag{4.12}$$

For each $f\in\mathscr{F}$, let $\bar\nu_m$ be the measure having $\bar f_m$ as a density with respect to ν_0 where, for every $f\in\mathscr{F}$, $\bar f_m$ is defined as follows.

$$\bar f_m(x) := \begin{cases} 1 & \text{if } x\in J_1, \\ \frac{\nu(J_j)}{\nu_0(J_j)} & \text{if } x\in J_j, \quad j=2,\dots,m. \end{cases} \tag{4.13}$$

Furthermore, define

$$\bar{\mathscr{P}}_n^{\nu_0} = \bigg(D, \mathscr{D}_{T_n}, \bigg\{P_{T_n}^{(\gamma^{\bar\nu_m-\nu_0},0,\bar\nu_m)} : \frac{d\bar\nu_m}{d\nu_0}\in\mathscr{F}\bigg\}\bigg). \tag{4.14}$$

4.4.1 Proof of Theorem 4.2.5

We begin by a series of lemmas that will be needed in the proof. Before doing so, let us underline the scheme of the proof. We recall that the goal is to prove that estimating $f = \frac{d\nu}{d\nu_0}$ from the continuous observation of a Lévy process $(X_t)_{t\in[0,T_n]}$ without Gaussian part and having Lévy measure ν is asymptotically equivalent to estimating f from the Gaussian white noise model:

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n g(t)}}dW_t, \quad g = \frac{d\nu_0}{d\mathrm{Leb}}, \quad t \in I.$$

Also, recall the definition of $\hat{\nu}_m$ given in (4.5) and read $\mathscr{P}_1 \overset{\Delta}{\Longleftrightarrow} \mathscr{P}_2$ as $\mathscr{P}_1$ is asymptotically equivalent to $\mathscr{P}_2$. Then, we can outline the proof in the following way.

- Step 1: $P_{T_n}^{(\gamma^{\nu-\nu_0},0,\nu)} \overset{\Delta}{\Longleftrightarrow} P_{T_n}^{(\gamma^{\hat{\nu}_m-\nu_0},0,\hat{\nu}_m)}$;
- Step 2: $P_{T_n}^{(\gamma^{\hat{\nu}_m-\nu_0},0,\hat{\nu}_m)} \overset{\Delta}{\Longleftrightarrow} \bigotimes_{j=2}^m \mathscr{P}(T_n\nu(J_j))$ (Poisson approximation).

  Here $\bigotimes_{j=2}^m \mathscr{P}(T_n\nu(J_j))$ represents a statistical model associated with the observation of $m-1$ independent Poisson r.v. of parameters $T_n\nu(J_j)$;
- Step 3: $\bigotimes_{j=2}^m \mathscr{P}(T_n\nu(J_j)) \overset{\Delta}{\Longleftrightarrow} \bigotimes_{j=2}^m \mathscr{N}(2\sqrt{T_n\nu(J_j)},1)$ (Gaussian approximation);
- Step 4: $\bigotimes_{j=2}^m \mathscr{N}(2\sqrt{T_n\nu(J_j)},1) \overset{\Delta}{\Longleftrightarrow} (y_t)_{t\in I}$.

Lemmas 4.4.1–4.4.3 are the key ingredients of Step 2.

Lemma 4.4.1. *Let $\bar{\mathscr{P}}_n^{\nu_0}$ and $\mathscr{L}_m$ be the statistical models defined in* (4.14) *and* (4.11), *respectively. Under the Assumption (H2) we have:*

$$\Delta(\bar{\mathscr{P}}_n^{\nu_0}, \mathscr{L}_m) = 0, \text{ for all } m.$$

Proof. Denote by $\bar{\mathbb{N}} = \mathbb{N}\cup\{\infty\}$ and consider the statistics $S : (D, \mathscr{D}_{T_n}) \to \big(\bar{\mathbb{N}}^{m-1}, \mathcal{P}(\bar{\mathbb{N}}^{m-1})\big)$ defined by

$$S(x) = \Big(N_{T_n}^{x;\,2},\dots,N_{T_n}^{x;\,m}\Big) \quad \text{with} \quad N_{T_n}^{x;\,j} = \sum_{r\le T_n} \mathbb{I}_{J_j}(\Delta x_r). \tag{4.15}$$

An application of Theorem 4.5.17 to $P_{T_n}^{(\gamma^{\bar{\nu}_m-\nu_0},0,\bar{\nu}_m)}$ and $P_{T_n}^{(0,0,\nu_0)}$, yields

$$\frac{dP_{T_n}^{(\gamma^{\bar{\nu}_m-\nu_0},0,\bar{\nu}_m)}}{dP_{T_n}^{(0,0,\nu_0)}}(x) = \exp\bigg(\sum_{j=2}^m \bigg(\ln\Big(\frac{\nu(J_j)}{\nu_0(J_j)}\Big)\bigg) N_{T_n}^{x;j} - T_n\int_I (\bar{f}_m(y)-1)\nu_0(dy)\bigg).$$

Hence, by means of the Fisher factorization theorem, we conclude that S is a sufficient statistics for $\bar{\mathscr{P}}_n^{\nu_0}$. Furthermore, under $P_{T_n}^{(\gamma^{\bar{\nu}_m-\nu_0},0,\bar{\nu}_m)}$, the random variables $N_{T_n}^{x;j}$ have Poisson distributions Q_j^f with means $T_n\nu(J_j)$. Then, by means of Property 4.5.12, we get $\Delta(\bar{\mathscr{P}}_n^{\nu_0},\mathscr{L}_m)=0,$ for all m. $\square$

Let us denote by $\hat{Q}_j^f$ the law of a Poisson random variable with mean $T_n\int_{J_j}\hat{f}_m(y)\nu_0(dy)$ and let $\hat{\mathscr{L}}_m$ be the statistical model associated with the family of probabilities $\{\bigotimes_{j=2}^m\hat{Q}_j^f : f\in\mathscr{F}\}$.

Lemma 4.4.2.

$$\Delta(\mathscr{L}_m,\hat{\mathscr{L}}_m)\leq\sup_{f\in\mathscr{F}}\sqrt{\frac{T_n}{\kappa}\int_{I\setminus[0,\varepsilon_m]}\big(f(y)-\hat{f}_m(y)\big)^2\nu_0(dy)}.$$

Proof. By means of Facts 4.5.7–4.5.9, we get:

$$\begin{aligned}\Delta(\mathscr{L}_m,\hat{\mathscr{L}}_m)&\leq\sup_{f\in\mathscr{F}}H\bigg(\bigotimes_{j=2}^m Q_j^f,\bigotimes_{j=2}^m\hat{Q}_j^f\bigg)\\&\leq\sup_{f\in\mathscr{F}}\sqrt{\sum_{j=2}^m 2H^2(Q_j^f,\hat{Q}_j^f)}\\&=\sup_{f\in\mathscr{F}}\sqrt{2}\sqrt{\sum_{j=2}^m\bigg(1-\exp\bigg(-\frac{T_n}{2}\bigg[\sqrt{\int_{J_j}\hat{f}_m(y)\nu_0(dy)}-\sqrt{\int_{J_j}f(y)\nu_0(dy)}\bigg]^2\bigg)\bigg)}.\end{aligned}$$

By making use of the fact that $1-e^{-x}\leq x$ for all $x\geq 0$ and the equality $\sqrt{a}-\sqrt{b}=\frac{a-b}{\sqrt{a}+\sqrt{b}}$ combined with the lower bound $f\geq\kappa$ (that also implies $\hat{f}_m\geq\kappa$) and finally the Cauchy-Schwarz inequality, we obtain:

$$\begin{aligned}&1-\exp\bigg(-\frac{T_n}{2}\bigg[\sqrt{\int_{J_j}\hat{f}_m(y)\nu_0(dy)}-\sqrt{\int_{J_j}f(y)\nu_0(dy)}\bigg]^2\bigg)\\&\leq\frac{T_n}{2}\bigg[\sqrt{\int_{J_j}\hat{f}_m(y)\nu_0(dy)}-\sqrt{\int_{J_j}f(y)\nu_0(dy)}\bigg]^2\\&\leq\frac{T_n}{2}\frac{\Big(\int_{J_j}(f(y)-\hat{f}_m(y))\nu_0(dy)\Big)^2}{\kappa\nu_0(J_j)}\\&\leq\frac{T_n}{2\kappa}\int_{J_j}\big(f(y)-\hat{f}_m(y)\big)^2\nu_0(dy).\end{aligned}$$

Hence,

$$H\bigg(\bigotimes_{j=2}^m Q_j^f,\bigotimes_{j=2}^m \hat{Q}_j^f\bigg)\leq \sqrt{\frac{T_n}{\kappa}\int_{I\setminus[0,\varepsilon_m]}\big(f(y)-\hat{f}_m(y)\big)^2\nu_0(dy)}.$$

□

Lemma 4.4.3. *Let $\hat{\nu}_m$ and $\bar{\nu}_m$ the Lévy measures defined as in* (4.5) *and* (4.13), *respectively. For every $f\in\mathscr{F}$, there exists a Markov kernel K such that*

$$KP_{T_n}^{(\gamma^{\bar{\nu}_m-\nu_0},0,\bar{\nu}_m)}=P_{T_n}^{(\gamma^{\hat{\nu}_m-\nu_0},0,\hat{\nu}_m)}.$$

Proof. By construction, $\bar{\nu}_m$ and $\hat{\nu}_m$ coincide on $[0,\varepsilon_m]$. Let us denote by $\bar{\nu}_m^{\textnormal{res}}$ and $\hat{\nu}_m^{\textnormal{res}}$ the restriction on $I\setminus[0,\varepsilon_m]$ of $\bar{\nu}_m$ and $\hat{\nu}_m$ respectively, then it is enough to prove: $KP_{T_n}^{(\gamma^{\bar{\nu}_m^{\textnormal{res}}-\nu_0},0,\bar{\nu}_m^{\textnormal{res}})}=P_{T_n}^{(\gamma^{\hat{\nu}_m^{\textnormal{res}}-\nu_0},0,\hat{\nu}_m^{\textnormal{res}})}$. First of all, let us observe that the kernel M:

$$M(x,A)=\sum_{j=2}^m\mathbb{I}_{J_j}(x)\int_A V_j(y)\nu_0(dy),\quad x\in I\setminus[0,\varepsilon_m],\quad A\in\mathscr{B}(I\setminus[0,\varepsilon_m])$$

is defined in such a way that $M\bar{\nu}_m^{\textnormal{res}}=\hat{\nu}_m^{\textnormal{res}}$. Indeed, for all $A\in\mathscr{B}(I\setminus[0,\varepsilon_m])$,

$$\begin{aligned}M\bar{\nu}_m^{\textnormal{res}}(A)&=\sum_{j=2}^m\int_{J_j}M(x,A)\bar{\nu}_m^{\textnormal{res}}(dx)=\sum_{j=2}^m\int_{J_j}\bigg(\int_A V_j(y)\nu_0(dy)\bigg)\bar{\nu}_m^{\textnormal{res}}(dx)\\&=\sum_{j=2}^m\bigg(\int_A V_j(y)\nu_0(dy)\bigg)\nu(J_j)=\int_A\hat{f}_m(y)\nu_0(dy)=\hat{\nu}_m^{\textnormal{res}}(A).\end{aligned}\tag{4.16}$$

Observe that $(\gamma^{\bar{\nu}_m^{\textnormal{res}}-\nu_0},0,\bar{\nu}_m^{\textnormal{res}})$ and $(\gamma^{\hat{\nu}_m^{\textnormal{res}}-\nu_0},0,\hat{\nu}_m^{\textnormal{res}})$ are Lévy triplets associated with compound Poisson processes since $\bar{\nu}_m^{\textnormal{res}}$ and $\hat{\nu}_m^{\textnormal{res}}$ are finite Lévy measures. The Markov kernel K interchanging the laws of the Lévy processes is constructed explicitly in the case of compound Poisson processes. Indeed if $\bar{X}$ is the compound Poisson process having Lévy measure $\bar{\nu}_m^{\textnormal{res}}$, then $\bar{X}_t=\sum_{i=1}^{N_t}\bar{Y}_i$, where N_t is a Poisson process of intensity $\iota_m:=\bar{\nu}_m^{\textnormal{res}}(I\setminus[0,\varepsilon_m])$ and the $\bar{Y}_i$ are i.i.d. random variables with probability law $\frac{1}{\iota_m}\bar{\nu}_m^{\textnormal{res}}$. Moreover, given a trajectory of $\bar{X}$, both the trajectory $(n_t)_{t\in[0,T_n]}$ of the Poisson process $(N_t)_{t\in[0,T_n]}$ and the realizations $\bar{y}_i$ of $\bar{Y}_i$, $i=1,\dots,n_{T_n}$ are uniquely determined. This allows us to construct n_{T_n} i.i.d. random variables $\hat{Y}_i$ as follows: For every realization $\bar{y}_i$ of $\bar{Y}_i$, we define the realization $\hat{y}_i$ of $\hat{Y}_i$ by throwing it according to the probability law $M(\bar{y}_i,\cdot)$. Hence, thanks to (4.16), $(\hat{Y}_i)_i$ are i.i.d. random variables with probability law $\frac{1}{\iota_m}\hat{\nu}_m^{\textnormal{res}}$. The desired Markov kernel K (defined on the Skorokhod space) is then given by:

$$K:(\bar{X}_t)_{t\in[0,T_n]}\longmapsto\bigg(\hat{X}_t:=\sum_{i=1}^{N_t}\hat{Y}_i\bigg)_{t\in[0,T_n]}.$$

Finally, observe that, since

$$\iota_m = \int_{I\setminus[0,\varepsilon_m]} \bar{f}_m(y)\nu_0(dy) = \int_{I\setminus[0,\varepsilon_m]} f(y)\nu_0(dy) = \int_{I\setminus[0,\varepsilon_m]} \hat{f}_m(y)\nu_0(dy),$$

$(\hat{X}_t)_{t\in[0,T_n]}$ is a compound Poisson process with Lévy measure $\hat{\nu}_m^{\text{res}}$. $\square$

Let us now state two lemmas needed to understand Step 4.

Lemma 4.4.4. *Denote by $\mathscr{W}_m^{\#}$ the statistical model associated with the continuous observation of a trajectory from the Gaussian white noise:*

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n}\sqrt{g(t)}}dW_t, \quad t \in I\setminus[0,\varepsilon_m].$$

Then, according with the notation introduced in Section 4.2.1 and at the beginning of Section 4.4, we have

$$\Delta(\mathscr{N}_m, \mathscr{W}_m^{\#}) \leq 2\sqrt{T_n}\sup_{f\in\mathscr{F}}\big(A_m(f) + B_m(f)\big).$$

Proof. As a preliminary remark observe that $\mathscr{W}_m^{\#}$ is equivalent to the model that observes a trajectory from:

$$d\bar{y}_t = \sqrt{f(t)}g(t)dt + \frac{\sqrt{g(t)}}{2\sqrt{T_n}}dW_t, \quad t \in I\setminus[0,\varepsilon_m].$$

Let us denote by $\bar{Y}_j$ the increments of the process $(\bar{y}_t)$ over the intervals J_j, $j = 2,\dots,m$, i.e.

$$\bar{Y}_j := \bar{y}_{v_j} - \bar{y}_{v_{j-1}} \sim \mathscr{N}\bigg(\int_{J_j}\sqrt{f(y)}\nu_0(dy), \frac{\nu_0(J_j)}{4T_n}\bigg)$$

and denote by $\bar{\mathscr{N}_m}$ the statistical model associated with the distributions of these increments. As an intermediate result, we will prove that

$$\Delta(\mathscr{N}_m, \bar{\mathscr{N}_m}) \leq 2\sqrt{T_n}\sup_{f\in\mathscr{F}} B_m(f), \quad \text{for all m.} \tag{4.17}$$

To that aim, remark that the experiment $\bar{\mathscr{N}_m}$ is equivalent to observing $m-1$ independent Gaussian random variables of means $\frac{2\sqrt{T_n}}{\sqrt{\nu_0(J_j)}}\int_{J_j}\sqrt{f(y)}\nu_0(dy)$, $j = 2,\dots,m$ and variances identically 1, name this last experiment $\mathscr{N}_m^{\#}$. Hence, using also Property 4.5.6, Facts 4.5.7 and 4.5.10 we get:

$$\Delta(\mathscr{N}_m, \bar{\mathscr{N}_m}) \leq \Delta(\mathscr{N}_m, \mathscr{N}_m^{\#}) \leq \sqrt{\sum_{j=2}^m \bigg(\frac{2\sqrt{T_n}}{\sqrt{\nu_0(J_j)}}\int_{J_j}\sqrt{f(y)}\nu_0(dy) - 2\sqrt{T_n\nu(J_j)}\bigg)^2}.$$

Since it is clear that $\delta(\mathscr{W}_m^\#, \bar{\mathscr{N}_m}) = 0$, in order to bound $\Delta(\mathscr{N}_m, \mathscr{W}_m^\#)$ it is enough to bound $\delta(\bar{\mathscr{N}_m}, \mathscr{W}_m^\#)$. Using similar ideas as in Carter (2002) Section 8.2, we define a new stochastic process as:

$$Y_t^* = \sum_{j=2}^m \bar{Y}_j \int_{\varepsilon_m}^t V_j(y)\nu_0(dy) + \frac{1}{2\sqrt{T_n}} \sum_{j=2}^m \sqrt{\nu_0(J_j)} B_j(t), \quad t \in I \setminus [0, \varepsilon_m],$$

where the $(B_j(t))$ are independent centered Gaussian processes independent of (W_t) and with variances

$$\mathrm{Var}(B_j(t)) = \int_{\varepsilon_m}^t V_j(y)\nu_0(dy) - \bigg(\int_{\varepsilon_m}^t V_j(y)\nu_0(dy)\bigg)^2.$$

These processes can be constructed from a standard Brownian bridge $\{B(s), s \in [0,1]\}$, independent of (W_t), via

$$B_i(t) = B\bigg(\int_{\varepsilon_m}^t V_i(y)\nu_0(dy)\bigg).$$

By construction, (Y_t^*) is a Gaussian process with mean and variance given by, respectively:

$$\begin{aligned}
\mathbb{E}[Y_t^*] &= \sum_{j=2}^m \mathbb{E}[\bar{Y}_j] \int_{\varepsilon_m}^t V_j(y)\nu_0(dy) = \sum_{j=2}^m \bigg(\int_{J_j} \sqrt{f(y)}\nu_0(dy)\bigg) \int_{\varepsilon_m}^t V_j(y)\nu_0(dy),\\
\mathrm{Var}[Y_t^*] &= \sum_{j=2}^m \mathrm{Var}[\bar{Y}_j] \bigg(\int_{\varepsilon_m}^t V_j(y)\nu_0(dy)\bigg)^2 + \frac{1}{4T_n} \sum_{j=2}^m \nu_0(J_j)\mathrm{Var}(B_j(t))\\
&= \frac{1}{4T_n} \int_{\varepsilon_m}^t \sum_{j=2}^m \nu_0(J_j) V_j(y)\nu_0(dy) = \frac{1}{4T_n} \int_{\varepsilon_m}^t \nu_0(dy) = \frac{\nu_0([\varepsilon_m, t])}{4T_n}.
\end{aligned}$$

One can compute in the same way the covariance of (Y_t^*) finding that

$$\mathrm{Cov}(Y_s^*, Y_t^*) = \frac{\nu_0([\varepsilon_m, s])}{4T_n}, \ \forall s \le t.$$

We can then deduce that

$$Y_t^* = \int_{\varepsilon_m}^t \widehat{\sqrt{f}}_m(y)\nu_0(dy) + \int_{\varepsilon_m}^t \frac{\sqrt{g(s)}}{2\sqrt{T_n}} dW_s^*, \quad t \in I \setminus [0, \varepsilon_m],$$

where (W_t^*) is a standard Brownian motion and

$$\widehat{\sqrt{f}}_m(x) := \sum_{j=2}^m \bigg(\int_{J_j} \sqrt{f(y)}\nu_0(dy)\bigg) V_j(x).$$

Applying Fact 4.5.11, we get that the total variation distance between the process $(Y_t^*)_{t\in I\setminus[0,\varepsilon_m]}$ constructed from the random variables $\bar{Y}_j$, $j=2,\dots,m$ and the Gaussian process $(\bar{y}_t)_{t\in I\setminus[0,\varepsilon_m]}$ is bounded by

$$\sqrt{4T_n\int_{I\setminus[0,\varepsilon_m]}\big(\widehat{\sqrt{f}}_m-\sqrt{f(y)}\big)^2\nu_0(dy)},$$

which gives the term in $A_m(f)$. □

Lemma 4.4.5. *In accordance with the notation of Lemma 4.4.4, we have:*

$$\Delta(\mathscr{W}_m^{\#},\mathscr{W}_n^{\nu_0})=O\bigg(\sup_{f\in\mathscr{F}}\sqrt{T_n\int_0^{\varepsilon_m}\big(\sqrt{f(t)}-1\big)^2\nu_0(dt)}\bigg). \tag{4.18}$$

Proof. Clearly $\delta(\mathscr{W}_n^{\nu_0},\mathscr{W}_m^{\#})=0$. To show that $\delta(\mathscr{W}_m^{\#},\mathscr{W}_n^{\nu_0})\to 0$, let us consider a Markov kernel $K^{\#}$ from $C(I\setminus[0,\varepsilon_m])$ to $C(I)$ defined as follows: Introduce a Gaussian process, $(B_t^m)_{t\in[0,\varepsilon_m]}$ with mean equal to t and covariance

$$\mathrm{Cov}(B_s^m,B_t^m)=\int_0^{\varepsilon_m}\frac{1}{4T_ng(s)}\mathbb{I}_{[0,s]\cap[0,t]}(z)dz.$$

In particular,

$$\mathrm{Var}(B_t^m)=\int_0^t\frac{1}{4T_ng(s)}ds.$$

Consider it as a process on the whole of I by defining $B_t^m=B_{\varepsilon_m}^m$ $\forall t>\varepsilon_m$. Let ω_t be a trajectory in $C(I\setminus[0,\varepsilon_m])$, which again we constantly extend to a trajectory on the whole of I. Then, we define $K^{\#}$ by sending the trajectory ω_t to the trajectory $\omega_t+B_t^m$. If we define $\tilde{\mathbb{W}}_n$ as the law induced on $C(I)$ by

$$d\tilde{y}_t=h(t)dt+\frac{dW_t}{2\sqrt{T_ng(t)}},\quad t\in I,\quad h(t)=\begin{cases}1 & t\in[0,\varepsilon_m]\\ \sqrt{f(t)} & t\in I\setminus[0,\varepsilon_m],\end{cases}$$

then $K^{\#}\mathbb{W}_n^f|_{I\setminus[0,\varepsilon_m]}=\tilde{\mathbb{W}}_n$, where $\mathbb{W}_n^f$ is defined as in (4.8). By means of Fact 4.5.11 we deduce (4.18). □

Proof of Theorem 4.2.5. The proof of the theorem follows by combining the previous lemmas together:

- Step 1: Let us denote by $\hat{\mathscr{P}}_{n,m}^{\nu_0}$ the statistical model associated with the family of probabilities $\big(P_{T_n}^{(\gamma^{\hat{\nu}_m-\nu_0},0,\hat{\nu}_m)}:\frac{d\nu}{d\nu_0}\in\mathscr{F}\big)$. Thanks to Property 4.5.6, Fact 4.5.7 and Theorem 4.5.18 we have that

$$\Delta(\mathscr{P}_n^{\nu_0},\hat{\mathscr{P}}_{n,m}^{\nu_0})\le\sqrt{\frac{T_n}{2}}\sup_{f\in\mathscr{F}}H(f,\hat{f}_m).$$

- Step 2: On the one hand, thanks to Lemma 4.4.1, one has that the statistical model associated with the family of probability $(P_{T_n}^{(\gamma^{\bar\nu_m-\nu_0},0,\bar\nu_m)}:\frac{d\nu}{d\nu_0}\in\mathscr{F})$ is equivalent to $\mathscr{L}_m$. By means of Lemma 4.4.2 we can bound $\Delta(\mathscr{L}_m,\hat{\mathscr{L}}_m)$. On the other hand it is easy to see that $\delta(\hat{\mathscr{P}}_{n,m}^{\nu_0},\hat{\mathscr{L}}_m)=0$. Indeed, it is enough to consider the statistics
$$S:x\mapsto\Bigg(\sum_{r\leq T_n}\mathbb{I}_{J_2}(\Delta x_r),\dots,\sum_{r\leq T_n}\mathbb{I}_{J_m}(\Delta x_r)\Bigg)$$
since the law of the random variable $\sum_{r\leq T_n}\mathbb{I}_{J_j}(\Delta x_r)$ under $P_{T_n}^{(\gamma^{\hat\nu_m-\nu_0},0,\hat\nu_m)}$ is Poisson of parameter $T_n\int_{J_j}\hat f_m(y)\nu_0(dy)$ for all $j=2,\dots,m$. Finally, Lemmas 4.4.1 and 4.4.3 allow us to conclude that $\delta(\mathscr{L}_m,\hat{\mathscr{P}}_{n,m}^{\nu_0})=0$. Collecting all the pieces together, we get
$$\Delta(\hat{\mathscr{P}}_{n,m}^{\nu_0},\mathscr{L}_m)\leq\sup_{f\in\mathscr{F}}\sqrt{\frac{T_n}{\kappa}\int_{I\setminus[0,\varepsilon_m]}\big(f(y)-\hat f_m(y)\big)^2\nu_0(dy)}.$$

- Step 3: Applying Theorem 4.5.14 and Fact 4.5.8 we can pass from the Poisson approximation given by $\mathscr{L}_m$ to a Gaussian one obtaining
$$\Delta(\mathscr{L}_m,\mathscr{N}_m)=C\sup_{f\in\mathscr{F}}\sqrt{\sum_{j=2}^m\frac{2}{T_n\nu(J_j)}}\leq C\sqrt{\sum_{j=2}^m\frac{2\kappa}{T_n\nu_0(J_j)}}=C\sqrt{\frac{(m-1)2\kappa}{T_n\mu_m}}.$$

- Step 4: Finally, Lemmas 4.4.4 and 4.4.5 allow us to conclude that:
$$\begin{aligned}\Delta(\mathscr{P}_n^{\nu_0},\mathscr{W}_n^{\nu_0})&=O\Big(\sqrt{T_n}\sup_{f\in\mathscr{F}}\big(A_m(f)+B_m(f)+C_m\big)\Big)\\&\quad+O\Big(\sqrt{T_n}\sup_{f\in\mathscr{F}}\sqrt{\int_{I\setminus[0,\varepsilon_m]}\big(f(y)-\hat f_m(y)\big)^2\nu_0(dy)}+\sqrt{\frac{m}{T_n\mu_m}}\Big).\end{aligned}$$

□

4.4.2 Proof of Theorem 4.2.6

Again, before stating some technical lemmas, let us highlight the main ideas of the proof. We recall that the goal is to prove that estimating $f=\frac{d\nu}{d\nu_0}$ from the discrete observations $(X_{t_i})_{i=0}^n$ of a Lévy process without Gaussian component and having Lévy measure ν is asymptotically equivalent to estimating f from the Gaussian white noise model
$$dy_t=\sqrt{f(t)}dt+\frac{1}{2\sqrt{T_ng(t)}}dW_t,\quad g=\frac{d\nu_0}{d\mathrm{Leb}},\quad t\in I.$$

Reading $\mathscr{P}_1 \stackrel{\Delta}{\Longleftrightarrow} \mathscr{P}_2$ as $\mathscr{P}_1$ is asymptotically equivalent to $\mathscr{P}_2$, we have:

- Step 1. Clearly $(X_{t_i})_{i=0}^n \stackrel{\Delta}{\Longleftrightarrow} (X_{t_i} - X_{t_{i-1}})_{i=1}^n$. Moreover, $(X_{t_i} - X_{t_{i-1}})_i \stackrel{\Delta}{\Longleftrightarrow} (\epsilon_i Y_i)$ where (ϵ_i) are i.i.d Bernoulli r.v. with parameter $\alpha = \iota_m \Delta_n e^{-\iota_m \Delta_n}$, $\iota_m := \int_{I\setminus[0,\varepsilon_m]} f(y)\nu_0(dy)$ and $(Y_i)_i$ are i.i.d. r.v. independent of $(\epsilon_i)_{i=1}^n$ and of density $\frac{f}{\iota_m}$ with respect to $\nu_{0|_{I\setminus[0,\varepsilon_m]}}$;

- Step 2. $(\epsilon_i Y_i)_i \stackrel{\Delta}{\Longleftrightarrow} \mathcal{M}(n;(\gamma_j)_{j=1}^m)$, where $\mathcal{M}(n;(\gamma_j)_{j=1}^m)$ is a multinomial distribution with $\gamma_1 = 1-\alpha$ and $\gamma_i := \alpha\nu(J_i)$ $i = 2,\dots,m$;

- Step 3. Gaussian approximation: $\mathcal{M}(n;(\gamma_1,\dots\gamma_m)) \stackrel{\Delta}{\Longleftrightarrow} \bigotimes_{j=2}^m \mathscr{N}(2\sqrt{T_n\nu(J_j)},1)$;

- Step 4. $\bigotimes_{j=2}^m \mathscr{N}(2\sqrt{T_n\nu(J_j)},1) \stackrel{\Delta}{\Longleftrightarrow} (y_t)_{t\in I}$.

Lemma 4.4.6. *Let ν_i, $i = 1,2$, be Lévy measures such that $\nu_1 \ll \nu_2$ and $b_1 - b_2 = \int_{|y|\le 1} y(\nu_1-\nu_2)(dy) < \infty$. Then, for all $0 < t < \infty$, we have:*

$$\Big\|Q_t^{(b_1,0,\mu_1)} - Q_t^{(b_2,0,\mu_2)}\Big\|_{TV} \le \sqrt{\frac{t}{2}}H(\nu_1,\nu_2).$$

Proof. For all given t, let K_t be the Markov kernel defined as $K_t(\omega,A) := \mathbb{I}_A(\omega_t)$, $\forall\ A \in \mathscr{B}(\mathbb{R})$, $\forall\ \omega \in D$. Then we have:

$$\begin{aligned}\big\|Q_t^{(b_1,0,\nu_1)} - Q_t^{(b_2,0,\nu_2)}\big\|_{TV} &= \big\|K_t P_t^{(b_1,0,\nu_1)} - K_t P_t^{(b_2,0,\nu_2)}\big\|_{TV}\\ &\le \big\|P_t^{(b_1,0,\nu_1)} - P_t^{(b_2,0,\nu_2)}\big\|_{TV}\\ &\le \sqrt{\frac{t}{2}}H(\nu_1,\nu_2),\end{aligned}$$

where we have used that Markov kernels reduce the total variation distance and Theorem 4.5.18. ☐

Lemma 4.4.7. *Let $(P_i)_{i=1}^n$, $(Y_i)_{i=1}^n$ and $(\epsilon_i)_{i=1}^n$ be samples of, respectively, Poisson random variables $\mathscr{P}(\lambda_i)$, random variables with common distribution and Bernoulli random variables of parameters $\lambda_i e^{-\lambda_i}$, which are all independent. Let us denote by $Q_{(Y_i,P_i)}$ (resp. $Q_{(Y_i,\epsilon_i)}$) the law of $\sum_{j=1}^{P_i} Y_j$ (resp., $\epsilon_i Y_i$). Then:*

$$\Big\|\bigotimes_{i=1}^n Q_{(Y_i,P_i)} - \bigotimes_{i=1}^n Q_{(Y_i,\epsilon_i)}\Big\|_{TV} \le 2\sqrt{\sum_{i=1}^n \lambda_i^2}. \tag{4.19}$$

The proof of this Lemma can be found in Mariucci (2015b), Section 2.1.

Lemma 4.4.8. *Let f_m^{tr} be the truncated function defined as follows:*

$$f_m^{\mathrm{tr}}(x) = \begin{cases} 1 & \text{if } x \in [0, \varepsilon_m] \\ f(x) & \text{otherwise} \end{cases}$$

and let ν_m^{tr} (resp. ν_m^{res}) be the Lévy measure having f_m^{tr} (resp. $f|_{I \setminus [0,\varepsilon_m]}$) as a density with respect to ν_0. Denote by $\mathscr{Q}_n^{\mathrm{tr},\nu_0}$ the statistical model associated with the family of probabilities $\Big(\bigotimes_{i=1}^n Q_{t_i - t_{i-1}}^{(\gamma^{\nu_m^{\mathrm{tr}}-\nu_0},0,\nu_m^{\mathrm{tr}})} : \frac{d\nu_m^{\mathrm{tr}}}{d\nu_0} \in \mathscr{F}\Big)$ and by $\mathscr{Q}_n^{\mathrm{res},\nu_0}$ the model associated with the family of probabilities $\Big(\bigotimes_{i=1}^n Q_{t_i - t_{i-1}}^{(\gamma^{\nu_m^{\mathrm{res}}-\nu_0},0,\nu_m^{\mathrm{res}})} : \frac{d\nu_m^{\mathrm{res}}}{d\nu_0} \in \mathscr{F}\Big)$. Then:

$$\Delta(\mathscr{Q}_n^{\mathrm{tr},\nu_0}, \mathscr{Q}_n^{\mathrm{res},\nu_0}) = 0.$$

Proof. Let us start by proving that $\delta(\mathscr{Q}_n^{\mathrm{tr},\nu_0}, \mathscr{Q}_n^{\mathrm{res},\nu_0}) = 0$. For that, let us consider two independent Lévy processes, X^{tr} and X^0, of Lévy triplets given by $\big(\gamma^{\nu_m^{\mathrm{tr}}-\nu_0}, 0, \nu_m^{\mathrm{tr}} - \nu_0\big)$ and $\big(0, 0, \nu_0|_{[0,\varepsilon_m]}\big)$, respectively. Then it is clear (using the *Lévy-Khintchine formula*) that the random variable $X_t^{\mathrm{tr}} - X_t^0$ is a randomization of X_t^{tr} (since the law of X_t^0 does not depend on ν) having law $Q_t^{(\gamma^{\nu_m^{\mathrm{res}}-\nu_0},0,\nu_m^{\mathrm{res}})}$, for all $t \geq 0$. Similarly, one can prove that $\delta(\mathscr{Q}_n^{\mathrm{res},\nu_0}, \mathscr{Q}_n^{\mathrm{tr},\nu_0}) = 0$. □

Proof of Theorem 4.2.6. As a preliminary remark, observe that the model $\mathscr{Q}_n^{\nu_0}$ is equivalent to the one that observes the increments of $\big((x_t), P_{T_n}^{(\gamma^{\nu-\nu_0},0,\nu)}\big)$, that is, the model $\tilde{\mathscr{Q}}_n^{\nu_0}$ associated with the family of probabilities $\Big(\bigotimes_{i=1}^n Q_{t_i - t_{i-1}}^{(\gamma^{\nu-\nu_0},0,\nu)} : \frac{d\nu}{d\nu_0} \in \mathscr{F}\Big)$.

- Step 1: Facts 4.5.7–4.5.8 and Lemma 4.4.6 allow us to write

$$\Big\| \bigotimes_{i=1}^n Q_{\Delta_n}^{(\gamma^{\nu-\nu_0},0,\nu)} - \bigotimes_{i=1}^n Q_{\Delta_n}^{(\gamma^{\nu_m^{\mathrm{tr}}-\nu_0},0,\nu_m^{\mathrm{tr}})} \Big\|_{TV} \leq \sqrt{n \sqrt{\frac{\Delta_n}{2}} H(\nu, \nu_m^{\mathrm{tr}})}$$
$$= \sqrt{n \sqrt{\frac{\Delta_n}{2}} \sqrt{\int_0^{\varepsilon_m} \big(\sqrt{f(y)} - 1\big)^2 \nu_0(dy)}}.$$

  Using this bound together with Lemma 4.4.8 and the notation therein, we get $\Delta(\mathscr{Q}_n^{\nu_0}, \mathscr{Q}_n^{\mathrm{res},\nu_0}) \leq \sqrt{n \sqrt{\frac{\Delta_n}{2}} \sup_{f \in \mathscr{F}} H(f, f_m^{\mathrm{tr}})}$. Observe that ν_m^{res} is a finite Lévy measure, hence $\Big((x_t), P_{T_n}^{(\gamma^{\nu_m^{\mathrm{res}}},0,\nu_m^{\mathrm{res}})}\Big)$ is a compound Poisson process with intensity equal to $\iota_m := \int_{I \setminus [0,\varepsilon_m]} f(y) \nu_0(dy)$ and jumps size density $\frac{f(x)g(x)}{\iota_m}$, for all $x \in I \setminus [0, \varepsilon_m]$ (recall that we are assuming that ν_0 has a density g with respect to Lebesgue). In particular, this means that $Q_{\Delta_n}^{(\gamma^{\nu_m^{\mathrm{res}}},0,\nu_m^{\mathrm{res}})}$ can be seen as the law of the random variable

$\sum_{j=1}^{P_i} Y_j$ where P_i is a Poisson variable of mean $\iota_m\Delta_n$, independent from $(Y_i)_{i\geq 0}$, a sequence of i.i.d. random variables with density $\frac{fg}{\iota_m}\mathbb{I}_{I\setminus[0,\varepsilon_m]}$ with respect to Lebesgue. Remark also that ι_m is confined between $\kappa\nu_0\big(I\setminus[0,\varepsilon_m]\big)$ and $M\nu_0\big(I\setminus[0,\varepsilon_m]\big)$.

Let $(\epsilon_i)_{i\geq 0}$ be a sequence of i.i.d. Bernoulli variables, independent of $(Y_i)_{i\geq 0}$, with mean $\iota_m\Delta_n e^{-\iota_m\Delta_n}$. For $i=1,\dots,n$, denote by $Q_i^{\epsilon,f}$ the law of the variable $\epsilon_i Y_i$ and by $\mathscr{Q}_n^\epsilon$ the statistical model associated with the observations of the vector $(\epsilon_1Y_1,\dots,\epsilon_nY_n)$, i.e.

$$\mathscr{Q}_n^\epsilon=\bigg(I^n,\mathscr{B}(I^n),\bigg\{\bigotimes_{i=1}^n Q_i^{\epsilon,f}:f\in\mathscr{F}\bigg\}\bigg).$$

Furthermore, denote by $\tilde{Q}_i^f$ the law of $\sum_{j=1}^{P_i}Y_j$. Then an application of Lemma 4.4.7 yields:

$$\Big\|\bigotimes_{i=1}^n\tilde{Q}_i^f-\bigotimes_{i=1}^n Q_i^{\epsilon,f}\Big\|_{TV}\leq 2\iota_m\sqrt{n\Delta_n^2}\leq 2M\nu_0\big(I\setminus[0,\varepsilon_m]\big)\sqrt{n\Delta_n^2}.$$

Hence, we get:

$$\Delta(\mathscr{Q}_n^{\mathrm{res},\nu_0},\mathscr{Q}_n^\epsilon)=O\bigg(\nu_0\big(I\setminus[0,\varepsilon_m]\big)\sqrt{n\Delta_n^2}\bigg).\tag{4.20}$$

Here the O depends only on M.

- Step 2: Let us introduce the following random variables:

$$Z_1=\sum_{j=1}^n\mathbb{I}_{\{0\}}(\epsilon_jY_j);\quad Z_i=\sum_{j=1}^n\mathbb{I}_{J_i}(\epsilon_jY_j),\ i=2,\dots,m.$$

Observe that the law of the vector $(Z_1,\dots,Z_m)$ is multinomial $\mathcal{M}(n;\gamma_1,\dots,\gamma_m)$ where

$$\gamma_1=1-\iota_m\Delta_ne^{-\iota_m\Delta_n},\quad \gamma_i=\Delta_ne^{-\iota_m\Delta_n}\nu(J_i),\quad i=2,\dots,m.$$

Let us denote by $\mathcal{M}_n$ the statistical model associated with the observation of $(Z_1,\dots,Z_m)$. Clearly $\delta(\mathscr{Q}_n^\epsilon,\mathcal{M}_n)=0$. Indeed, $\mathcal{M}_n$ is the image experiment by the random variable $S:I^n\to\{1,\dots,n\}^m$ defined as

$$S(x_1,\dots,x_n)=\Big(\#\{j:x_j=0\};\#\big\{j:x_j\in J_2\big\};\dots;\#\big\{j:x_j\in J_m\big\}\Big),$$

where $\#A$ denotes the cardinal of the set A.

We shall now prove that $\delta(\mathcal{M}_n,\mathscr{Q}_n^\epsilon)\leq \sup_{f\in\mathscr{F}}\sqrt{n\Delta_n H^2(f,\hat f_m)}$. We start by defining a discrete random variable X^* concentrated at the points 0, x_i^*, $i=2,\dots,m$:

$$\mathbb{P}(X^*=y)=\begin{cases}\gamma_i & \text{if } y=x_i^*, \quad i=1,\dots,m,\\ 0 & \text{otherwise,}\end{cases}$$

with the convention $x_1^*=0$. It is easy to see that $\mathcal{M}_n$ is equivalent to the statistical model associated with n independent copies of X^*. Let us introduce the Markov kernel

$$K(x_i^*,A)=\begin{cases}\mathbb{I}_A(0) & \text{if } i=1,\\ \int_A V_i(x)\nu_0(dx) & \text{otherwise.}\end{cases}$$

Denote by P^* the law of the random variable X^* and by $Q_i^{\epsilon,\hat f}$ the law of a random variable $\epsilon_i\hat Y_i$ where ϵ_i is Bernoulli independent of $\hat Y_i$, with mean $\iota_m\Delta_n e^{-\iota_m\Delta_n}$ and $\hat Y_i$ has a density $\frac{\hat f_m g}{\iota_m}\mathbb{I}_{I\setminus[0,\varepsilon_m]}$ with respect to Lebesgue. The same computations as in Lemma 4.4.3 prove that $KP^*=Q_i^{\epsilon,\hat f}$. Hence, thanks to Remark 4.5.13, we get the equivalence between $\mathcal{M}_n$ and the statistical model associated with the observations of n independent copies of $\epsilon_i\hat Y_i$. In order to bound $\delta(\mathcal{M}_n,\mathscr{Q}_n^\epsilon)$ it is enough to bound the total variation distance between the probabilities $\bigotimes_{i=1}^n Q_i^{\epsilon,f}$ and $\bigotimes_{i=1}^n Q_i^{\epsilon,\hat f}$. Alternatively, we can bound the Hellinger distance between each of the $Q_i^{\epsilon,f}$ and $Q_i^{\epsilon,\hat f}$, thanks to Facts 4.5.7 and 4.5.8, which is:

$$\begin{aligned}\Big\|\bigotimes_{i=1}^n Q_i^{\epsilon,f}-\bigotimes_{i=1}^n Q_i^{\epsilon,\hat f}\Big\|_{TV} &\leq \sqrt{\sum_{i=1}^n H^2\big(Q_i^{\epsilon,f},Q_i^{\epsilon,\hat f}\big)}\\ &=\sqrt{\sum_{i=1}^n \frac{1-\gamma_1}{\iota}H^2(f,\hat f_m)}\leq\sqrt{n\Delta_n H^2(f,\hat f_m)}.\end{aligned}$$

It follows that

$$\delta(\mathcal{M}_n,\mathscr{Q}_n^\epsilon)\leq\sqrt{n\Delta_n}\sup_{f\in\mathscr{F}}H(f,\hat f_m).$$

- Step 3: Let us denote by $\mathcal{N}_m^*$ the statistical model associated with the observation of m independent Gaussian variables $\mathscr{N}(n\gamma_i,n\gamma_i)$, $i=1,\dots,m$. Very similar computations to those in Carter (2002) yield

$$\Delta(\mathcal{M}_n,\mathcal{N}_m^*)=O\Big(\frac{m\ln m}{\sqrt n}\Big).$$

In order to prove the asymptotic equivalence between $\mathcal{M}_n$ and $\mathcal{N}_m$ defined as in (4.12) we need to introduce some auxiliary statistical models. Let us denote by $\mathcal{A}_m$ the experiment obtained from $\mathcal{N}_m^*$ by disregarding the first component. Furthermore, let us denote by $\mathcal{N}_m^{\#}$ the experiment associated with the observation of $m-1$ independent Gaussian variables $\mathscr{N}(\sqrt{n\gamma_i}, \frac{1}{4})$, $i = 2, \dots, m$. First of all, let us prove that $\mathcal{N}_m^*$ and $\mathcal{A}_m$ are asymptotically equivalent. One direction is trivial. In the other direction, the Markov kernel is given by

$$K(x_2, \dots, x_m, A) = \mathbb{E}\big[\mathbb{I}_A(X, x_2, \dots, x_m)\big], \quad A \subset \mathbb{R}^m,$$

where X is a Gaussian random variable with mean and variance both equal to n. The image experiment of $\mathcal{A}_m$ through K is the statistical model associated with the observation of m independent Gaussian random variables of the form $\mathcal{N}(n, n) \otimes \bigotimes_{i=2}^m \mathcal{N}(n\gamma_i, n\gamma_i)$. The total variation distance can then be computed explicitly: It equals the distance between the first components, for which the general formula of Fact 4.5.10 can be used. We get the bound:

$$\Delta(\mathcal{N}_m^*, \mathcal{A}_m) \leq \sup_{f \in \mathscr{F}} \sqrt{\Big(2 + \frac{n}{2}\Big)(1 - \gamma_1)^2} = O\Big(\nu_0\big(I \setminus [0, \varepsilon_m]\big)\Delta_n \sqrt{n}\Big).$$

Moreover, using a result contained in Carter (2002), Section 7.2, one has that

$$\Delta(\mathcal{A}_m, \mathcal{N}_m^{\#}) = O\Big(\frac{m}{\sqrt{n}}\Big).$$

Finally, using Facts 4.5.7 and 4.5.10 we can write

$$\begin{aligned}\Delta(\mathcal{N}_m^{\#}, \mathcal{N}_m) &\leq \sqrt{2\sum_{i=2}^m \Big(\sqrt{T_n \nu(J_i)} - \sqrt{T_n \nu(J_i)\exp(-\iota_m \Delta_n)}\Big)^2} \\ &\leq \sqrt{2T_n \Delta_n^2 \iota_m^3} \leq \sqrt{2n\Delta_n^3 M^3 \big(\nu_0\big(I \setminus [0, \varepsilon_m]\big)\big)^3}.\end{aligned}$$

To sum up,

$$\Delta(\mathcal{M}_n, \mathcal{N}_m) = O\Big(\frac{m \ln m}{\sqrt{n}} + \sqrt{n\Delta_n^3 \big(\nu_0\big(I \setminus [0, \varepsilon_m]\big)\big)^3} + \nu_0\big(I \setminus [0, \varepsilon_m]\big)\Delta_n \sqrt{n}\Big),$$

with the O depending only on κ and M.

- Step 4: An application of Lemmas 4.4.4 and 4.4.5 yields

$$\Delta(\mathcal{N}_m, \mathscr{W}_n^{\nu_0}) \leq 2\sqrt{T_n} \sup_{f \in \mathscr{F}} \big(A_m(f) + B_m(f) + C_m(f)\big).$$

□

4.5 Proofs of the examples

The purpose of this section is to give detailed proofs of Examples 4.2.4 and Examples 4.3.1–4.3.3. As in Section 4.4 we suppose $I \subseteq \mathbb{R}_+$. We start by giving some bounds for the quantities $A_m(f)$, $B_m(f)$ and $L_2(f, \hat{f}_m)$, the L_2-distance between the restriction of f and $\hat{f}_m$ on $I \setminus [0, \varepsilon_m]$.

4.5.1 Bounds for $A_m(f)$, $B_m(f)$, $L_2(f, \hat{f}_m)$ when $\hat{f}_m$ is piecewise linear.

In this section we suppose f to be in $\mathscr{F}^I_{(\gamma,K,\kappa,M)}$ defined as in (4.3). We are going to assume that the V_j are given by triangular/trapezoidal functions as in (4.6). In particular, in this case $\hat{f}_m$ is piecewise linear.

Lemma 4.5.1. *Let $0 < \kappa < M$ be two constants and let f_i, $i = 1, 2$ be functions defined on an interval J and such that $\kappa \leq f_i \leq M$, $i = 1, 2$. Then, for any measure ν_0, we have:*

$$\begin{aligned} \frac{1}{4M} \int_J \big(f_1(x) - f_2(x)\big)^2 \nu_0(dx) &\leq \int_J \big(\sqrt{f_1(x)} - \sqrt{f_2(x)}\big)^2 \nu_0(dx) \\ &\leq \frac{1}{4\kappa} \int_J \big(f_1(x) - f_2(x)\big)^2 \nu_0(dx). \end{aligned}$$

Proof. This simply comes from the following inequalities:

$$\begin{aligned} \frac{1}{2\sqrt{M}} |f_1(x) - f_2(x)| &\leq \frac{|f_1(x) - f_2(x)|}{\sqrt{f_1(x)} + \sqrt{f_2(x)}} = \big|\sqrt{f_1(x)} - \sqrt{f_2(x)}\big| \\ &\leq \frac{1}{2\sqrt{\kappa}} |f_1(x) - f_2(x)|. \end{aligned}$$

□

Recall that x_i^* is chosen so that $\int_{J_i} (x - x_i^*) \nu_0(dx) = 0$. Consider the following Taylor expansions for $x \in J_i$:

$$f(x) = f(x_i^*) + f'(x_i^*)(x - x_i^*) + R_i(x); \quad \hat{f}_m(x) = \hat{f}_m(x_i^*) + \hat{f}'_m(x_i^*)(x - x_i^*),$$

where $\hat{f}_m(x_i^*) = \frac{\nu(J_i)}{\nu_0(J_i)}$ and $\hat{f}'_m(x_i^*)$ is the left or right derivative in x_i^* depending whether $x < x_i^*$ or $x > x_i^*$ (as $\hat{f}_m$ is piecewise linear, no rest is involved in its Taylor expansion).

Lemma 4.5.2. *The following estimates hold:*

$$\begin{aligned}|R_i(x)| &\leq K|\xi_i - x_i^*|^\gamma |x - x_i^*|;\\ \big|f(x_i^*) - \hat{f}_m(x_i^*)\big| &\leq \|R_i\|_{L_\infty(\nu_0)} \text{ for } i = 2,\dots,m-1;\\ \big|f(x) - \hat{f}_m(x)\big| &\leq \begin{cases} 2\|R_i\|_{L_\infty(\nu_0)} + K|x_i^* - \eta_i|^\gamma |x - x_i^*| & \text{if } x \in J_i,\ i = 3,\dots,m-1;\\ C|x - \tau_i| & \text{if } x \in J_i,\ i \in \{2, m\}.\end{cases}\end{aligned}$$

for some constant C and points $\xi_i \in J_i$, $\eta_i \in J_{i-1} \cup J_i \cup J_{i+1}$, $\tau_2 \in J_2 \cup J_3$ and $\tau_m \in J_{m-1} \cup J_m$.

Proof. By definition of R_i, we have

$$|R_i(x)| = \Big|\big(f'(\xi_i) - f'(x_i^*)\big)(x - x_i^*)\Big| \leq K|\xi_i - x_i^*|^\gamma |x - x_i^*|,$$

for some point $\xi_i \in J_i$. For the second inequality,

$$\begin{aligned}|f(x_i^*) - \hat{f}_m(x_i^*)| &= \frac{1}{\nu_0(J_i)}\Big|\int_{J_i}(f(x_i^*) - f(x))\nu_0(dx)\Big|\\ &= \frac{1}{\nu_0(J_i)}\Big|\int_{J_i} R_i(x)\nu_0(dx)\Big| \leq \|R_i\|_{L_\infty(\nu_0)},\end{aligned}$$

where in the first inequality we have used the defining property of x_i^*. For the third inequality, let us start by proving that for all $2 < i < m-1$, $\hat{f}_m'(x_i^*) = f'(\chi_i)$ for some $\chi_i \in J_i \cup J_{i+1}$ (here, we are considering right derivatives; for left ones, this would be $J_{i-1} \cup J_i$). To see that, take $x \in J_i \cap [x_i^*, x_{i+1}^*]$ and introduce the function $h(x) := f(x) - l(x)$ where

$$l(x) = \frac{x - x_i^*}{x_{i+1}^* - x_i^*}\big(\hat{f}_m(x_{i+1}^*) - \hat{f}_m(x_i^*)\big) + \hat{f}_m(x_i^*).$$

Then, using the fact that $\int_{J_i}(x - x_i^*)\nu_0(dx) = 0$ together with $\int_{J_{i+1}}(x - x_i^*)\nu_0(dx) = (x_{j+1}^* - x_j^*)\mu_m$, we get

$$\int_{J_i} h(x)\nu_0(dx) = 0 = \int_{J_{i+1}} h(x)\nu_0(dx).$$

In particular, by means of the mean theorem, one can conclude that there exist two points $p_i \in J_i$ and $p_{i+1} \in J_{i+1}$ such that

$$h(p_i) = \frac{\int_{J_i} h(x)\nu_0(dx)}{\nu_0(J_i)} = \frac{\int_{J_{i+1}} h(x)\nu_0(dx)}{\nu_0(J_{i+1})} = h(p_{i+1}).$$

As a consequence, we can deduce that there exists $\chi_i \in [p_i, p_{i+1}] \subseteq J_i \cup J_{i+1}$ such that $h'(\chi_i) = 0$, hence $f'(\chi_i) = l'(\chi_i) = \hat{f}_m'(x_i^*)$. When $2 < i < m-1$, the two Taylor expansions together with the fact that $\hat{f}_m'(x_i^*) = f'(\chi_i)$ for some $\chi_i \in J_i \cup J_{i+1}$, give

$$\begin{aligned}|f(x) - \hat{f}_m(x)| &\leq |f(x_i^*) - \hat{f}_m(x_i^*)| + |R_i(x)| + K|x_i^* - \chi_i|^\gamma |x - x_i^*|\\ &\leq 2\|R_i\|_{L_\infty(\nu_0)} + K|x_i^* - \chi_i|^\gamma |x - x_i^*|\end{aligned}$$

whenever $x \in J_i$ and $x > x_i^*$ (the case $x < x_i^*$ is handled similarly using the left derivative of $\hat{f}_m$ and $\xi_i \in J_{i-1} \cup J_i$). For the remaining cases, consider for example $i = 2$. Then $\hat{f}_m(x)$ is bounded by the minimum and the maximum of f on $J_2 \cup J_3$, hence $\hat{f}_m(x) = f(\tau)$ for some $\tau \in J_2 \cup J_3$. Since f' is bounded by $C = 2M + K$, one has $|f(x) - \hat{f}_m(x)| \leq C|x - \tau|$. $\square$

Lemma 4.5.3. *With the same notations as in Lemma 4.5.2, the estimates for $A_m^2(f)$, $B_m^2(f)$ and $L_2(f, \hat{f}_m)^2$ are as follows:*

$$\begin{aligned}L_2(f, \hat{f}_m)^2 &\leq \frac{1}{4\kappa}\bigg(\sum_{i=3}^m \int_{J_i} \Big(2\|R_i\|_{L_\infty(\nu_0)} + K|x_i^* - \eta_i|^\gamma |x - x_i^*|\Big)^2 \nu_0(dx)\\ &\qquad + C^2\Big(\int_{J_2} |x - \tau_2|^2 \nu_0(dx) + \int_{J_m} |x - \tau_m|^2 \nu_0(dx)\Big)\bigg).\\ A_m^2(f) &= L_2\big(\sqrt{f}, \widehat{\sqrt{f}}_m\big)^2 = O\Big(L_2(f, \hat{f}_m)^2\Big)\\ B_m^2(f) &= O\bigg(\sum_{i=2}^m \frac{1}{\sqrt{\kappa}} \nu_0(J_i)(2\sqrt{M} + 1)^2 \|R_i\|_{L_\infty(\nu_0)}^2\bigg).\end{aligned}$$

Proof. The L_2-bound is now a straightforward application of Lemmas 4.5.1 and 4.5.2. The one on $A_m(f)$ follows, since if $f \in \mathscr{F}^I_{(\gamma,K,\kappa,M)}$ then $\sqrt{f} \in \mathscr{F}^I_{(\gamma,\frac{K}{\sqrt{\kappa}},\sqrt{\kappa},\sqrt{M})}$. In order to bound $B_m^2(f)$ write it as:

$$B_m^2(f) = \sum_{j=1}^m \nu_0(J_j)\bigg(\frac{\int_{J_j}\sqrt{f(y)}\nu_0(dy)}{\nu_0(J_j)} - \sqrt{\frac{\nu(J_j)}{\nu_0(J_j)}}\bigg)^2 =: \sum_{j=1}^m \nu_0(J_j) E_j^2.$$

By the triangular inequality, let us bound E_j by $F_j + G_j$ where:

$$F_j = \bigg|\sqrt{\frac{\nu(J_j)}{\nu_0(J_j)}} - \sqrt{f(x_j^*)}\bigg| \quad \text{and} \quad G_j = \bigg|\sqrt{f(x_j^*)} - \frac{\int_{J_j}\sqrt{f(y)}\nu_0(dy)}{\nu_0(J_j)}\bigg|.$$

Using the same trick as in the proof of Lemma 4.5.1, we can bound:

$$F_j \leq 2\sqrt{M}\bigg|\frac{\int_{J_j}\big(f(x) - f(x_i^*)\big)\nu_0(dx)}{\nu_0(J_j)}\bigg| \leq 2\sqrt{M}\|R_j\|_{L_\infty(\nu_0)}.$$

On the other hand,

$$\begin{aligned}
G_j &= \frac{1}{\nu_0(J_j)}\bigg|\int_{J_j}\big(\sqrt{f(x_j^*)}-\sqrt{f(y)}\big)\nu_0(dy)\bigg| \\
&= \frac{1}{\nu_0(J_j)}\bigg|\int_{J_j}\bigg(\frac{f'(x_j^*)}{2\sqrt{f(x_j^*)}}(x-x_j^*)+\tilde{R}_j(y)\bigg)\nu_0(dy)\bigg| \leq \|\tilde{R}_j\|_{L_\infty(\nu_0)},
\end{aligned}$$

which has the same magnitude as $\frac{1}{\kappa}\|R_j\|_{L_\infty(\nu_0)}$.

□

Remark 4.5.4. Observe that when ν_0 is finite, there is no need for a special definition of $\hat{f}_m$ near 0, and all the estimates in Lemma 4.5.2 hold true replacing every occurrence of $i=2$ by $i=1$.

Remark 4.5.5. The same computations as in Lemmas 4.5.2 and 4.5.3 can be adapted to the general case where the V_j's (and hence $\hat{f}_m$) are not piecewise linear. In the general case, the Taylor expansion of $\hat{f}_m$ in x_i^* involves a rest as well, say $\hat{R}_i$, and one needs to bound this, as well.

4.5.2 Proofs of Examples 4.2.4

In the following, we collect the details of the proofs of Examples 4.2.4.

1. The finite case: $\nu_0 \equiv \mathrm{Leb}([0,1])$.

Remark that in the case where ν_0 if finite there are no convergence problems near zero and so we can consider the easier approximation of f:

$$\hat{f}_m(x) := \begin{cases} m\theta_1 & \text{if } x\in\big[0,x_1^*\big], \\ m^2\big[\theta_{j+1}(x-x_j^*)+\theta_j(x_{j+1}^*-x)\big] & \text{if } x\in(x_j^*,x_{j+1}^*] \quad j=1,\dots,m-1, \\ m\theta_m & \text{if } x\in(x_m^*,1] \end{cases}$$

where

$$x_j^*=\frac{2j-1}{2m},\quad J_j=\Big(\frac{j-1}{m},\frac{j}{m}\Big],\quad \theta_j=\int_{J_j}f(x)dx,\quad j=1,\dots,m.$$

In this case we take $\varepsilon_m=0$ and Conditions $(C2)$ and $(C2')$ coincide:

$$\lim_{n\to\infty} n\Delta_n \sup_{f\in\mathscr{F}}\Big(A_m^2(f)+B_m^2(f)\Big)=0.$$

Applying Lemma 4.5.3, we get

$$\sup_{f\in\mathscr{F}}\Big(L_2(f,\hat{f}_m)+A_m(f)+B_m(f)\Big)=O\big(m^{-\frac{3}{2}}+m^{-1-\gamma}\big);$$

(actually, each of the three terms on the left hand side has the same rate of convergence).

2. The finite variation case: $\frac{d\nu_0}{d\mathrm{Leb}}(x) = x^{-1}\mathbb{I}_{[0,1]}(x)$.

To prove that the standard choice of V_j described at the beginning of Examples 4.2.4 leads to $\displaystyle\int_{\varepsilon_m}^1 V_j(x)\frac{dx}{x} = 1$, it is enough to prove that this integral is independent of j, since in general $\displaystyle\int_{\varepsilon_m}^1 \sum_{j=2}^m V_j(x)\frac{dx}{x} = m-1$. To that aim observe that, for $j = 3,\dots,m-1$,

$$\mu_m\int_{\varepsilon_m}^1 V_j(x)\nu_0(dx) = \int_{x_{j-1}^*}^{x_j^*}\frac{x-x_{j-1}^*}{x_j^*-x_{j-1}^*}\frac{dx}{x} + \int_{x_j^*}^{x_{j+1}^*}\frac{x_{j+1}^*-x}{x_{j+1}^*-x_j^*}\frac{dx}{x}.$$

Let us show that the first addendum does not depend on j. We have

$$\int_{x_{j-1}^*}^{x_j^*}\frac{dx}{x_j^*-x_{j-1}^*} = 1 \quad \text{and} \quad -\frac{x_{j-1}^*}{x_j^*-x_{j-1}^*}\int_{x_{j-1}^*}^{x_j^*}\frac{dx}{x} = \frac{x_{j-1}^*}{x_j^*-x_{j-1}^*}\ln\Big(\frac{x_{j-1}^*}{x_j^*}\Big).$$

Since $x_j^* = \frac{v_j-v_{j-1}}{\mu_m}$ and $v_j = \varepsilon_m^{\frac{m-j}{m-1}}$, the quantities $\frac{x_j^*}{x_{j-1}^*}$ and, hence, $\frac{x_{j-1}^*}{x_j^*-x_{j-1}^*}$ do not depend on j. The second addendum and the trapezoidal functions V_2 and V_m are handled similarly. Thus, $\hat f_m$ can be chosen of the form

$$\hat f_m(x) := \begin{cases} 1 & \text{if } x\in\big[0,\varepsilon_m\big],\\ \frac{\nu(J_2)}{\mu_m} & \text{if } x\in\big(\varepsilon_m, x_2^*\big],\\ \frac{1}{x_{j+1}^*-x_j^*}\Big[\frac{\nu(J_{j+1})}{\mu_m}(x-x_j^*)+\frac{\nu(J_j)}{\mu_m}(x_{j+1}^*-x)\Big] & \text{if } x\in(x_j^*,x_{j+1}^*] \quad j=2,\dots,m-1,\\ \frac{\nu(J_m)}{\mu_m} & \text{if } x\in(x_m^*,1]. \end{cases}$$

A straightforward application of Lemmas 4.5.2 and 4.5.3 gives

$$\sqrt{\int_{\varepsilon_m}^1\Big(f(x)-\hat f_m(x)\Big)^2\nu_0(dx)} + A_m(f) + B_m(f) = O\bigg(\Big(\frac{\ln m}{m}\Big)^{\gamma+1}\sqrt{\ln(\varepsilon_m^{-1})}\bigg),$$

as announced.

3. The infinite variation, non-compactly supported case: $\frac{d\nu_0}{d\mathrm{Leb}}(x) = x^{-2}\mathbb{I}_{\mathbb{R}_+}(x)$. Recall that we want to prove that

$$L_2(f,\hat f_m)^2 + A_m^2(f) + B_m^2(f) = O\bigg(\frac{H(m)^{3+4\gamma}}{(\varepsilon_m m)^{2\gamma}} + \sup_{x\geq H(m)}\frac{f(x)^2}{H(m)}\bigg),$$

for any given sequence $H(m)$ going to infinity as $m\to\infty$.

Let us start by addressing the problem that the triangular/trapezoidal choice for V_j is not doable. Introduce the following notation: $V_j = \overset{\triangle}{V_j} + A_j$, $j = 2, \dots, m$, where the $\overset{\triangle}{V_j}$'s are triangular/trapezoidal function similar to those in (4.6). The difference is that here, since x_m^* is not defined, $\overset{\triangle}{V}_{m-1}$ is a trapezoid, linear between x_{m-2}^* and x_{m-1}^* and constantly equal to $\frac{1}{\mu_m}$ on $[x_{m-1}^*, v_{m-1}]$ and $\overset{\triangle}{V}_m$ is supported on $[v_{m-1}, \infty)$, where it is constantly equal to $\frac{1}{\mu_m}$. Each A_j is chosen so that:

1. It is supported on $[x_{j-1}^*, x_{j+1}^*]$ (unless $j = 2$, $j = m-1$ or $j = m$; in the first case the support is $[x_2^*, x_3^*]$, in the second one it is $[x_{m-2}^*, x_{m-1}^*]$, and $A_m \equiv 0$);
2. A_j coincides with $-A_{j-1}$ on $[x_{j-1}^*, x_j^*]$, $j = 3, \dots, m-1$ (so that $\sum V_j \equiv \frac{1}{\mu_m}$) and its first derivative is bounded (in absolute value) by $\frac{1}{\mu_m(x_j^* - x_{j-1}^*)}$ (so that V_j is non-negative and bounded by $\frac{1}{\mu_m}$);
3. A_j vanishes, along with its first derivatives, on x_{j-1}^*, x_j^* and x_{j+1}^*.

We claim that these conditions are sufficient to assure that $\hat{f}_m$ converges to f quickly enough. First of all, by Remark 4.5.5, we observe that, to have a good bound on $L_2(f, \hat{f}_m)$, the crucial property of $\hat{f}_m$ is that its first right (resp. left) derivative has to be equal to $\frac{1}{\mu_m(x_{j+1}^* - x_j^*)}$ (resp. $\frac{1}{\mu_m(x_j^* - x_{j-1}^*)}$) and its second derivative has to be small enough (for example, so that the rest $\hat{R}_j$ is as small as the rest R_j of f already appearing in Lemma 4.5.2).

The (say) left derivatives in x_j^* of $\hat{f}_m$ are given by

$$\hat{f}_m'(x_j^*) = \big(\overset{\triangle}{V_j}'(x_j^*) + A_j'(x_j^*)\big)\big(\nu(J_j) - \nu(J_{j-1})\big); \quad \hat{f}_m''(x_j^*) = A_j''(x_j^*)\big(\nu(J_j) - \nu(J_{j-1})\big).$$

Then, in order to bound $|\hat{f}_m''(x_j^*)|$ it is enough to bound $|A_j''(x_j^*)|$ because:

$$\big|\hat{f}_m''(x_j^*)\big| \leq |A_j''(x_j^*)| \Big| \int_{J_j} f(x)\frac{dx}{x^2} - \int_{J_{j-1}} f(x)\frac{dx}{x^2} \Big| \leq |A_j''(x_j^*)| \sup_{x \in I} |f'(x)| (\ell_j + \ell_{j-1})\mu_m,$$

where ℓ_j is the Lebesgue measure of J_j.

We are thus left to show that we can choose the A_j's satisfying points 1-3, with a small enough second derivative, and such that $\int_I V_j(x)\frac{dx}{x^2} = 1$. To make computations easier, we will make the following explicit choice:

$$A_j(x) = b_j(x - x_j^*)^2(x - x_{j-1}^*)^2 \quad \forall x \in [x_{j-1}^*, x_j^*),$$

for some b_j depending only on j and m (the definitions on $[x_j^*, x_{j+1}^*)$ are uniquely determined by the condition $A_j + A_{j+1} \equiv 0$ there).

Define $j_{\max}$ as the index such that $H(m) \in J_{j_{\max}}$; it is straightforward to check that

$$j_{\max} \sim m - \frac{\varepsilon_m(m-1)}{H(m)}; \quad x^*_{m-k} = \varepsilon_m(m-1)\log\Big(1+\frac{1}{k}\Big), \quad k=1,\dots,m-2.$$

One may compute the following Taylor expansions:

$$\int_{x^*_{m-k-1}}^{x^*_{m-k}} \overset{\triangle}{V}_{m-k}(x)\nu_0(dx) = \frac{1}{2} - \frac{1}{6k} + \frac{5}{24k^2} + O\Big(\frac{1}{k^3}\Big);$$
$$\int_{x^*_{m-k}}^{x^*_{m-k+1}} \overset{\triangle}{V}_{m-k}(x)\nu_0(dx) = \frac{1}{2} + \frac{1}{6k} + \frac{1}{24k^2} + O\Big(\frac{1}{k^3}\Big).$$

In particular, for $m \gg 0$ and $m-k \leq j_{\max}$, so that also $k \gg 0$, all the integrals $\int_{x^*_{j-1}}^{x^*_{j+1}} \overset{\triangle}{V}_j(x)\nu_0(dx)$ are bigger than 1 (it is immediate to see that the same is true for $\overset{\triangle}{V}_2$, as well). From now on we will fix a $k \geq \frac{\varepsilon_m m}{H(m)}$ and let $j = m-k$.

Summing together the conditions $\int_I V_i(x)\nu_0(dx) = 1\ \forall i > j$ and noticing that the function $\sum_{i=j}^m V_i$ is constantly equal to $\frac{1}{\mu_m}$ on $[x_j^*, \infty)$ we have:

$$\begin{aligned}\int_{x^*_{j-1}}^{x^*_j} A_j(x)\nu_0(dx) &= m-j+1-\frac{1}{\mu_m}\nu_0([x^*_j,\infty)) - \int_{x^*_{j-1}}^{x^*_j} \overset{\triangle}{V}_j(x)\nu_0(dx)\\ &= k+1-\frac{1}{\log(1+\frac{1}{k})} - \frac{1}{2} + \frac{1}{6k} + O\Big(\frac{1}{k^2}\Big) = \frac{1}{4k} + O\Big(\frac{1}{k^2}\Big)\end{aligned}$$

Our choice of A_j allows us to compute this integral explicitly:

$$\int_{x^*_{j-1}}^{x^*_j} b_j(x-x^*_{j-1})^2(x-x^*_j)^2\frac{dx}{x^2} = b_j\big(\varepsilon_m(m-1)\big)^3\Big(\frac{2}{3}\frac{1}{k^4} + O\Big(\frac{1}{k^5}\Big)\Big).$$

In particular one gets that asymptotically

$$b_j \sim \frac{1}{(\varepsilon_m(m-1))^3}\frac{3}{2}k^4\frac{1}{4k} \sim \Big(\frac{k}{\varepsilon_m m}\Big)^3.$$

This immediately allows us to bound the first order derivative of A_j as asked in point 2: Indeed, it is bounded above by $2b_j\ell^3_{j-1}$ where ℓ_{j-1} is again the length of J_{j-1}, namely $\ell_j = \frac{\varepsilon_m(m-1)}{k(k+1)} \sim \frac{\varepsilon_m m}{k^2}$. It follows that for m big enough:

$$\sup_{x\in I}|A_j'(x)| \leq \frac{1}{k^3} \ll \frac{1}{\mu_m(x^*_j - x^*_{j-1})} \sim \Big(\frac{k}{\varepsilon_m m}\Big)^2.$$

The second order derivative of $A_j(x)$ can be easily computed to be bounded by $4b_j\ell_j^2$. Also remark that the conditions that $|f|$ is bounded by M and that f' is Hölder, say

$|f'(x) - f'(y)| \leq K|x-y|^\gamma$, together give a uniform L_∞ bound of $|f'|$ by $2M + K$. Summing up, we obtain:

$$|\hat{f}_m''(x_j^*)| \lesssim b_j \ell_m^3 \mu_m \sim \frac{1}{k^3 \varepsilon_m m}$$

(here and in the following we use the symbol $\lesssim$ to stress that we work up to constants and to higher order terms). The leading term of the rest $\hat{R}_j$ of the Taylor expansion of $\hat{f}_m$ near x_j^* is

$$\hat{f}_m''(x_j^*)|x - x_j^*|^2 \sim |f_m''(x_j^*)|\ell_j^2 \sim \frac{\varepsilon_m m}{k^7}.$$

Using Lemmas 4.5.2 and 4.5.3 (taking into consideration Remark 4.5.5) we obtain

$$\begin{aligned}\int_{\varepsilon_m}^\infty |f(x) - \hat{f}_m(x)|^2 \nu_0(dx) &\lesssim \sum_{j=2}^{j_{\max}} \int_{J_j} |f(x) - \hat{f}_m(x)|^2 \nu_0(dx) + \int_{H(m)}^\infty |f(x) - \hat{f}_m(x)|^2 \nu_0(dx) \\ &\lesssim \sum_{k=\frac{\varepsilon_m m}{H(m)}}^{m} \mu_m \bigg(\frac{(\varepsilon_m m)^{2+2\gamma}}{k^{4+4\gamma}} + \frac{(\varepsilon_m m)^2}{k^{14}} \bigg) + \frac{1}{H(m)} \sup_{x \geq H(m)} f(x)^2 \\ &\lesssim \bigg(\frac{H(m)^{3+4\gamma}}{(\varepsilon_m m)^{2+2\gamma}} + \frac{H(m)^{13}}{(\varepsilon_m m)^{10}} \bigg) + \frac{1}{H(m)}.\end{aligned} \tag{4.21}$$

It is easy to see that, since $0 < \gamma \leq 1$, as soon as the first term converges, it does so more slowly than the second one. Thus, an optimal choice for $H(m)$ is given by $\sqrt{\varepsilon_m m}$, that gives a rate of convergence:

$$L_2(f, \hat{f}_m)^2 \lesssim \frac{1}{\sqrt{\varepsilon_m m}}.$$

This directly gives a bound on $H(f, \hat{f}_m)$. Also, the bound on the term $A_m(f)$, which is $L_2(\sqrt{f}, \widehat{\sqrt{f}}_m)^2$, follows as well, since $f \in \mathscr{F}^I_{(\gamma,K,\kappa,M)}$ implies $\sqrt{f} \in \mathscr{F}^I_{(\gamma,\frac{K}{\sqrt{\kappa}},\sqrt{\kappa},\sqrt{M})}$. Finally, the term $B_m^2(f)$ contributes with the same rates as those in (4.21): Using Lemma 4.5.3,

$$\begin{aligned}B_m^2(f) &\lesssim \sum_{j=2}^{\lceil m - \frac{\varepsilon_m(m-1)}{H(m)} \rceil} \nu_0(J_j) \|R_j\|_{L_\infty}^2 + \nu_0([H(m), \infty)) \\ &\lesssim \mu_m \sum_{k=\frac{\varepsilon_m(m-1)}{H(m)}}^{m} \Big(\frac{\varepsilon_m m}{k^2} \Big)^{2+2\gamma} + \frac{1}{H(m)} \\ &\lesssim \frac{H(m)^{3+4\gamma}}{(\varepsilon_m m)^{2+2\gamma}} + \frac{1}{H(m)}.\end{aligned}$$

4.5.3 Proof of Example 4.3.1

In this case, since $\varepsilon_m = 0$, the proofs of Theorems 4.2.5 and 4.2.6 simplify and give better estimates near zero, namely:

$$\begin{aligned}
\Delta(\mathscr{P}_{n,FV}^{\mathrm{Leb}}, \mathscr{W}_n^{\nu_0}) &\leq C_1\bigg(\sqrt{T_n}\sup_{f\in\mathscr{F}}\Big(A_m(f)+B_m(f)+L_2(f,\hat{f}_m)\Big)+\sqrt{\frac{m^2}{T_n}}\bigg)\\
\Delta(\mathscr{Q}_{n,FV}^{\mathrm{Leb}}, \mathscr{W}_n^{\nu_0}) &\leq C_2\bigg(\sqrt{n\Delta_n^2}+\frac{m\ln m}{\sqrt{n}}+\sqrt{T_n}\sup_{f\in\mathscr{F}}\Big(A_m(f)+B_m(f)+H\big(f,\hat{f}_m\big)\Big)\bigg),
\end{aligned} \tag{4.22}$$

where C_1, C_2 depend only on κ, M and

$$A_m(f)=\sqrt{\int_0^1\Big(\widehat{\sqrt{f}}_m(y)-\sqrt{f(y)}\Big)^2dy},\quad B_m(f)=\sum_{j=1}^m\bigg(\sqrt{m}\int_{J_j}\sqrt{f(y)}dy-\sqrt{\theta_j}\bigg)^2.$$

As a consequence we get:

$$\Delta(\mathscr{P}_{n,FV}^{\mathrm{Leb}}, \mathscr{W}_n^{\nu_0}) \leq O\bigg(\sqrt{T_n}(m^{-\frac{3}{2}}+m^{-1-\gamma})+\sqrt{m^2T_n^{-1}}\bigg).$$

To get the bounds in the statement of Example 4.3.1 the optimal choices are $m_n = T_n^{\frac{1}{2+\gamma}}$ when $\gamma\leq\frac{1}{2}$ and $m_n = T_n^{\frac{2}{5}}$ otherwise. Concerning the discrete model, we have:

$$\Delta(\mathscr{Q}_{n,FV}^{\mathrm{Leb}}, \mathscr{W}_n^{\nu_0}) \leq O\bigg(\sqrt{n\Delta_n^2}+\frac{m\ln m}{\sqrt{n}}+\sqrt{n\Delta_n}\big(m^{-\frac{3}{2}}+m^{-1-\gamma}\big)\bigg).$$

There are four possible scenarios: If $\gamma > \frac{1}{2}$ and $\Delta_n = n^{-\beta}$ with $\frac{1}{2}<\beta<\frac{3}{4}$ (resp. $\beta\geq\frac{3}{4}$) then the optimal choice is $m_n = n^{1-\beta}$ (resp. $m_n = n^{\frac{2-\beta}{5}}$).

If $\gamma\geq\frac{1}{2}$ and $\Delta_n = n^{-\beta}$ with $\frac{1}{2}<\beta<\frac{2+2\gamma}{3+2\gamma}$ (resp. $\beta\geq\frac{2+2\gamma}{3+2\gamma}$) then the optimal choice is $m_n = n^{\frac{2-\beta}{4+2\gamma}}$ (resp. $m_n = n^{1-\beta}$).

4.5.4 Proof of Example 4.3.2

As in Examples 4.2.4, we let $\varepsilon_m = m^{-1-\alpha}$ and consider the standard triangular/trapezoidal V_j's. In particular, $\hat{f}_m$ will be piecewise linear. Condition (C2') is satisfied and we have $C_m(f) = O(\varepsilon_m)$. This bound, combined with the one obtained in (4.7), allows us to conclude that an upper bound for the rate of convergence of $\Delta(\mathscr{Q}_{n,FV}^{\nu_0}, \mathscr{W}_n^{\nu_0})$ is given by:

$$\Delta(\mathscr{Q}_{n,FV}^{\nu_0}, \mathscr{W}_n^{\nu_0}) \leq C\bigg(\sqrt{\sqrt{n^2\Delta_n}\varepsilon_m+\sqrt{n\Delta_n}\Big(\frac{\ln(\varepsilon_m^{-1})}{m}\Big)^2}+\frac{m\ln m}{\sqrt{n}}+\sqrt{n\Delta_n^2}\ln(\varepsilon_m^{-1})\bigg),$$

where C is a constant only depending on the bound on $\lambda > 0$.

The sequences ε_m and m can be chosen arbitrarily to optimize the rate of convergence. It is clear from the expression above that, if we take $\varepsilon_m = m^{-1-\alpha}$ with $\alpha > 0$, bigger values of α reduce the first term $\sqrt{\sqrt{n^2\Delta_n}\varepsilon_m}$, while changing the other terms only by constants. It can be seen that taking $\alpha \geq 15$ is enough to make the first term negligible with respect to the others. In that case, and under the assumption $\Delta_n = n^{-\beta}$, the optimal choice for m is $m = n^\delta$ with $\delta = \frac{5-4\beta}{14}$. In that case, the global rate of convergence is

$$\Delta(\mathscr{Q}_{n,FV}^{\nu_0}, \mathscr{W}_n^{\nu_0}) = \begin{cases} O\big(n^{\frac{1}{2}-\beta}\ln n\big) & \text{if } \frac{1}{2} < \beta \leq \frac{9}{10} \\ O\big(n^{-\frac{1+2\beta}{7}}\ln n\big) & \text{if } \frac{9}{10} < \beta < 1. \end{cases}$$

In the same way one can find

$$\Delta(\mathscr{P}_{n,FV}^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\bigg(\sqrt{n\Delta_n}\Big(\frac{\ln m}{m}\Big)^2\sqrt{\ln(\varepsilon_m^{-1})} + \sqrt{\frac{m^2}{n\Delta_n\ln(\varepsilon_m)}} + \sqrt{n\Delta_n}\varepsilon_m\bigg).$$

As above, we can freely choose ε_m and m (in a possibly different way from above). Again, as soon as $\varepsilon_m = m^{-1-\alpha}$ with $\alpha \geq 1$ the third term plays no role, so that we can choose $\varepsilon_m = m^{-2}$. Letting $\Delta_n = n^{-\beta}$, $0 < \beta < 1$, and $m = n^\delta$, an optimal choice is $\delta = \frac{1-\beta}{3}$, giving

$$\Delta(\mathscr{P}_{n,FV}^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\Big(n^{\frac{\beta-1}{6}}\big(\ln n\big)^{\frac{5}{2}}\Big) = O\Big(T_n^{-\frac{1}{6}}\big(\ln T_n\big)^{\frac{5}{2}}\Big).$$

4.5.5 Proof of Example 4.3.3

Using the computations in (4.21), combined with $\big(f(y) - \hat{f}_m(y)\big)^2 \leq 4\exp(-2\lambda_0 y^3) \leq 4\exp(-2\lambda_0 H(m)^3)$ for all $y \geq H(m)$, we obtain:

$$\begin{aligned} \int_{\varepsilon_m}^\infty \big|f(x) - \hat{f}_m(x)\big|^2\nu_0(dx) &\lesssim \frac{H(m)^7}{(\varepsilon_m m)^4} + \int_{H(m)}^\infty \big|f(x) - \hat{f}_m(x)\big|^2\nu_0(dx) \\ &\lesssim \frac{H(m)^7}{(\varepsilon_m m)^4} + \frac{e^{-2\lambda_0 H(m)^3}}{H(m)}. \end{aligned}$$

As in Example 4.2.4, this bounds directly $H^2(f,\hat f_m)$ and $A_m^2(f)$. Again, the first part of the integral appearing in $B_m^2(f)$ is asymptotically smaller than the one appearing above:

$$\begin{aligned}B_m^2(f)&=\sum_{j=1}^m\bigg(\frac{1}{\sqrt{\mu_m}}\int_{J_j}\sqrt{f}\nu_0-\sqrt{\int_{J_j}f(x)\nu_0(dx)}\bigg)^2\\&\lesssim\frac{H(m)^7}{(\varepsilon_m m)^4}+\sum_{k=1}^{\frac{\varepsilon_m m}{H(m)}}\bigg(\frac{1}{\sqrt{\mu_m}}\int_{J_{m-k}}\sqrt{f}\nu_0-\sqrt{\int_{J_{m-k}}f(x)\nu_0(dx)}\bigg)^2\\&\lesssim\frac{H(m)^7}{(\varepsilon_m m)^4}+\frac{e^{-\lambda_0 H(m)^3}}{H(m)}.\end{aligned}$$

As above, for the last inequality we have bounded f in each J_{m-k}, $k\leq\frac{\varepsilon_m m}{H(m)}$, with $\exp(-\lambda_0 H(m)^3)$. Thus the global rate of convergence of $L_2(f,\hat f_m)^2+A_m^2(f)+B_m^2(f)$ is $\frac{H(m)^7}{(\varepsilon_m m)^4}+\frac{e^{-\lambda_0 H(m)^3}}{H(m)}$.

Concerning $C_m(f)$, we have $C_m^2(f)=\int_0^{\varepsilon_m}\frac{(\sqrt{f(x)}-1)^2}{x^2}dx\lesssim\varepsilon_m^5$. To write the global rate of convergence of the Le Cam distance in the discrete setting we make the choice $H(m)=\sqrt[3]{\frac{\eta}{\lambda_0}\ln m}$, for some constant η, and obtain:

$$\Delta(\mathscr{Q}_n^{\nu_0},\mathscr{W}_n^{\nu_0})=O\bigg(\frac{\sqrt{n}\Delta_n}{\varepsilon_m}+\frac{m\ln m}{\sqrt{n}}+\sqrt{n\Delta_n}\Big(\frac{(\ln m)^{\frac76}}{(\varepsilon_m m)^2}+\frac{m^{-\frac{\eta}{2}}}{\sqrt[3]{\ln m}}\Big)+\sqrt[4]{n^2\Delta_n\varepsilon_m^5}\bigg).$$

Letting $\Delta_n=n^{-\beta}$, $\varepsilon_m=n^{-\alpha}$ and $m=n^\delta$, optimal choices give $\alpha=\frac{\beta}{3}$ and $\delta=\frac13+\frac{\beta}{18}$. We can also take $\eta=2$ to get a final rate of convergence:

$$\Delta(\mathscr{Q}_n^{\nu_0},\mathscr{W}_n^{\nu_0})=\begin{cases}O\big(n^{\frac12-\frac23\beta}\big)&\text{if } \frac34<\beta<\frac{12}{13}\\O\big(n^{-\frac16+\frac{\beta}{18}}(\ln n)^{\frac76}\big)&\text{if } \frac{12}{13}\leq\beta<1.\end{cases}$$

In the continuous setting, we have

$$\Delta(\mathscr{P}_n^{\nu_0},\mathscr{W}_n^{\nu_0})=O\bigg(\sqrt{n\Delta_n}\Big(\frac{(\ln m)^{\frac76}}{(\varepsilon_m m)^2}+\frac{m^{-\frac{\eta}{2}}}{\sqrt[3]{\ln m}}+\varepsilon_m^{\frac52}\Big)+\sqrt{\frac{\varepsilon_m m^2}{n\Delta_n}}\bigg).$$

Using $T_n=n\Delta_n$, $\varepsilon_m=T_n^{-\alpha}$ and $m=T_n^\delta$, optimal choices are given by $\alpha=\frac{4}{17}$, $\delta=\frac{9}{17}$; choosing any $\eta\geq3$ we get the rate of convergence

$$\Delta(\mathscr{P}_n^{\nu_0},\mathscr{W}_n^{\nu_0})=O\big(T_n^{-\frac{3}{34}}(\ln T_n)^{\frac76}\big).$$

Background

Le Cam theory of statistical experiments

A *statistical model* or *experiment* is a triplet $\mathscr{P}_j = (\mathscr{X}_j, \mathscr{A}_j, \{P_{j,\theta}; \theta \in \Theta\})$ where $\{P_{j,\theta}; \theta \in \Theta\}$ is a family of probability distributions all defined on the same σ-field $\mathscr{A}_j$ over the *sample space* $\mathscr{X}_j$ and Θ is the *parameter space*. The *deficiency* $\delta(\mathscr{P}_1, \mathscr{P}_2)$ of $\mathscr{P}_1$ with respect to $\mathscr{P}_2$ quantifies "how much information we lose" by using $\mathscr{P}_1$ instead of $\mathscr{P}_2$ and it is defined as $\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_K \sup_{\theta\in\Theta} ||KP_{1,\theta} - P_{2,\theta}||_{TV}$, where TV stands for "total variation" and the infimum is taken over all "transitions" K (see Le Cam (1986), page 18). The general definition of transition is quite involved but, for our purposes, it is enough to know that Markov kernels are special cases of transitions. By $KP_{1,\theta}$ we mean the image measure of $P_{1,\theta}$ via the Markov kernel K, that is

$$KP_{1,\theta}(A) = \int_{\mathscr{X}_1} K(x, A)P_{1,\theta}(dx), \quad \forall A \in \mathscr{A}_2.$$

The experiment $K\mathscr{P}_1 = (\mathscr{X}_2, \mathscr{A}_2, \{KP_{1,\theta}; \theta \in \Theta\})$ is called a *randomization* of $\mathscr{P}_1$ by the Markov kernel K. When the kernel K is deterministic, that is $K(x, A) = \mathbb{I}_A S(x)$ for some random variable $S : (\mathscr{X}_1, \mathscr{A}_1) \to (\mathscr{X}_2, \mathscr{A}_2)$, the experiment $K\mathscr{P}_1$ is called the *image experiment by the random variable* S. The Le Cam distance is defined as the symmetrization of δ and it defines a pseudometric. When $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$ the two statistical models are said to be *equivalent*. Two sequences of statistical models $(\mathscr{P}_1^n)_{n\in\mathbb{N}}$ and $(\mathscr{P}_2^n)_{n\in\mathbb{N}}$ are called *asymptotically equivalent* if $\Delta(\mathscr{P}_1^n, \mathscr{P}_2^n)$ tends to zero as n goes to infinity. A very interesting feature of the Le Cam distance is that it can be also translated in terms of statistical decision theory. Let $\mathscr{D}$ be any (measurable) decision space and let $L : \Theta \times \mathscr{D} \mapsto [0, \infty)$ denote a loss function. Let $\|L\| = \sup_{(\theta,z)\in\Theta\times\mathscr{D}} L(\theta, z)$. Let π_i denote a (randomized) decision procedure in the i-th experiment. Denote by $R_i(\pi_i, L, \theta)$ the risk from using procedure π_i when L is the loss function and θ is the true value of the parameter. Then, an equivalent definition of the deficiency is:

$$\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_{\pi_1} \sup_{\pi_2} \sup_{\theta\in\Theta} \sup_{L:\|L\|=1} \big|R_1(\pi_1, L, \theta) - R_2(\pi_2, L, \theta)\big|.$$

Thus $\Delta(\mathscr{P}_1, \mathscr{P}_2) < \varepsilon$ means that for every procedure π_i in problem i there is a procedure π_j in problem j, $\{i, j\} = \{1, 2\}$, with risks differing by at most ε, uniformly over all bounded L and $\theta \in \Theta$. In particular, when minimax rates of convergence in a nonparametric estimation problem are obtained in one experiment, the same rates automatically hold in any asymptotically equivalent experiment. There is more: When explicit transformations from one experiment to another are obtained, statistical procedures can be carried over from one experiment to the other one.

There are various techniques to bound the Le Cam distance. We report below only the properties that are useful for our purposes. For the proofs see, e.g., Le Cam (1986); Strasser (1985).

Property 4.5.6. *Let $\mathscr{P}_j = (\mathscr{X}, \mathscr{A}, \{P_{j,\theta}; \theta \in \Theta\})$, $j = 1, 2$, be two statistical models having the same sample space and define $\Delta_0(\mathscr{P}_1, \mathscr{P}_2) := \sup_{\theta\in\Theta} \|P_{1,\theta} - P_{2,\theta}\|_{TV}$. Then, $\Delta(\mathscr{P}_1, \mathscr{P}_2) \leq \Delta_0(\mathscr{P}_1, \mathscr{P}_2)$.*

In particular, Property 4.5.6 allows us to bound the Le Cam distance between statistical models sharing the same sample space by means of classical bounds for the total variation distance. To that aim, we collect below some useful results.

Fact 4.5.7. *Let P_1 and P_2 be two probability measures on $\mathscr{X}$, dominated by a common measure ξ, with densities $g_i = \frac{dP_i}{d\xi}$, $i = 1, 2$. Define*

$$L_1(P_1, P_2) = \int_{\mathscr{X}} |g_1(x) - g_2(x)|\xi(dx),$$
$$H(P_1, P_2) = \bigg(\int_{\mathscr{X}} \Big(\sqrt{g_1(x)} - \sqrt{g_2(x)}\Big)^2 \xi(dx)\bigg)^{1/2}.$$

Then,

$$\|P_1 - P_2\|_{TV} = \frac{1}{2} L_1(P_1, P_2) \leq H(P_1, P_2). \tag{4.23}$$

Fact 4.5.8. *Let P and Q be two product measures defined on the same sample space: $P = \otimes_{i=1}^n P_i$, $Q = \otimes_{i=1}^n Q_i$. Then*

$$H^2(P, Q) \leq \sum_{i=1}^n H^2(P_i, Q_i). \tag{4.24}$$

Fact 4.5.9. *Let P_i, $i = 1, 2$, be the law of a Poisson random variable with mean λ_i. Then*

$$H^2(P_1, P_2) = 1 - \exp\bigg(-\frac{1}{2}\Big(\sqrt{\lambda_1} - \sqrt{\lambda_2}\Big)^2\bigg).$$

Fact 4.5.10. *Let $Q_1 \sim \mathscr{N}(\mu_1, \sigma_1^2)$ and $Q_2 \sim \mathscr{N}(\mu_2, \sigma_2^2)$. Then*

$$\|Q_1 - Q_2\|_{TV} \leq \sqrt{2\bigg(1 - \frac{\sigma_1^2}{\sigma_2^2}\bigg)^2 + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}}.$$

Fact 4.5.11. *For $i=1,2$, let Q_i, $i=1,2$, be the law on $(C,\mathscr{C})$ of two Gaussian processes of the form*

$$X_t^i=\int_0^t h_i(s)ds+\int_0^t \sigma(s)dW_s,\ t\in[0,T]$$

where $h_i\in L_2(\mathbb{R})$ and $\sigma\in\mathbb{R}_{>0}$. Then:

$$L_1\big(Q_1,Q_2\big)\leq\sqrt{\int_0^T\frac{\big(h_1(y)-h_2(y)\big)^2}{\sigma^2(s)}ds}.$$

Property 4.5.12. *Let $\mathscr{P}_i=(\mathscr{X}_i,\mathscr{A}_i,\{P_{i,\theta},\theta\in\Theta\})$, $i=1,2$, be two statistical models. Let $S:\mathscr{X}_1\to\mathscr{X}_2$ be a sufficient statistics such that the distribution of S under $P_{1,\theta}$ is equal to $P_{2,\theta}$. Then $\Delta(\mathscr{P}_1,\mathscr{P}_2)=0$.*

Remark 4.5.13. Let P_i be a probability measure on $(E_i,\mathcal{E}_i)$ and K_i a Markov kernel on $(G_i,\mathcal{G}_i)$. One can then define a Markov kernel K on $(\prod_{i=1}^n E_i,\otimes_{i=1}^n\mathcal{G}_i)$ in the following way:

$$K(x_1,\dots,x_n;A_1\times\dots\times A_n):=\prod_{i=1}^n K_i(x_i,A_i),\quad \forall x_i\in E_i,\ \forall A_i\in\mathcal{G}_i.$$

Clearly $K\otimes_{i=1}^n P_i=\otimes_{i=1}^n K_iP_i$.

Finally, we recall the following result that allows us to bound the Le Cam distance between Poisson and Gaussian variables.

Theorem 4.5.14. *(See Brown et al. (2004a), Theorem 4) Let $\tilde{P}_\lambda$ be the law of a Poisson random variable $\tilde{X}_\lambda$ with mean λ. Furthermore, let P_λ^* be the law of a random variable Z_λ^* with Gaussian distribution $\mathscr{N}(2\sqrt{\lambda},1)$, and let $\tilde{U}$ be a uniform variable on $\big[-\frac{1}{2},\frac{1}{2}\big)$ independent of $\tilde{X}_\lambda$. Define*

$$\tilde{Z}_\lambda=2\mathrm{sgn}\big(\tilde{X}_\lambda+\tilde{U}\big)\sqrt{\big|\tilde{X}_\lambda+\tilde{U}\big|}. \tag{4.25}$$

Then, denoting by P_λ the law of $\tilde{Z}_\lambda$,

$$H^2\big(P_\lambda,P_\lambda^*\big)=O(\lambda^{-1}).$$

Remark 4.5.15. Thanks to Theorem 4.5.14, denoting by Λ a subset of $\mathbb{R}_{>0}$, by $\tilde{\mathscr{P}}$ (resp. $\mathscr{P}^*$) the statistical model associated with the family of probabilities $\{\tilde{P}_\lambda:\lambda\in\Lambda\}$ (resp. $\{P_\lambda^*:\lambda\in\Lambda\}$), we have

$$\Delta\big(\tilde{\mathscr{P}},\mathscr{P}^*\big)\leq\sup_{\lambda\in\Lambda}\frac{C}{\lambda},$$

for some constant C. Indeed, the correspondence associating $\tilde{Z}_\lambda$ to $\tilde{X}_\lambda$ defines a Markov kernel; conversely, associating to $\tilde{Z}_\lambda$ the closest integer to its square, defines a Markov kernel going in the other direction.

Lévy processes

Definition 4.5.16. A stochastic process $\{X_t : t \geq 0\}$ on $\mathbb{R}$ defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ is called a *Lévy process* if the following conditions are satisfied.

1. $X_0 = 0$ $\mathbb{P}$-a.s.
2. For any choice of $n \geq 1$ and $0 \leq t_0 < t_1 < \ldots < t_n$, random variables X_{t_0}, $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$are independent.
3. The distribution of $X_{s+t} - X_s$ does not depend on s.
4. There is $\Omega_0 \in \mathscr{A}$ with $\mathbb{P}(\Omega_0) = 1$ such that, for every $\omega \in \Omega_0$, $X_t(\omega)$ is right-continuous in $t \geq 0$ and has left limits in $t > 0$.
5. It is stochastically continuous.

Thanks to the *Lévy-Khintchine formula*, the characteristic function of any Lévy process $\{X_t\}$ can be expressed, for all u in $\mathbb{R}$, as:

$$\mathbb{E}\big[e^{iuX_t}\big] = \exp\bigg(- t\Big(iub - \frac{u^2\sigma^2}{2} - \int_{\mathbb{R}} (1 - e^{iuy} + iuy\mathbb{I}_{|y|\leq 1})\nu(dy)\Big)\bigg),$$

where $b, \sigma \in \mathbb{R}$ and ν is a measure on $\mathbb{R}$ satisfying

$$\nu(\{0\}) = 0 \text{ and } \int_{\mathbb{R}} (|y|^2 \wedge 1)\nu(dy) < \infty.$$

In the sequel we shall refer to (b, σ^2, ν) as the characteristic triplet of the process $\{X_t\}$ and ν will be called the *Lévy measure*. This data characterizes uniquely the law of the process $\{X_t\}$.

Let $D = D([0, \infty), \mathbb{R})$ be the space of mappings ω from $[0, \infty)$ into $\mathbb{R}$ that are right-continuous with left limits. Define the *canonical process* $x : D \to D$ by

$$\forall \omega \in D, \quad x_t(\omega) = \omega_t, \;\; \forall t \geq 0.$$

Let $\mathscr{D}_t$ and $\mathscr{D}$ be the σ-algebras generated by $\{x_s : 0 \leq s \leq t\}$ and $\{x_s : 0 \leq s < \infty\}$, respectively (here, we use the same notations as in Sato (1999)).

By the condition (4) above, any Lévy process on $\mathbb{R}$ induces a probability measure P on $(D, \mathscr{D})$. Thus $\{X_t\}$ on the probability space $(D, \mathscr{D}, P)$ is identical in law with the original Lévy process. By saying that $(\{x_t\}, P)$ is a Lévy process, we mean that $\{x_t : t \geq 0\}$ is a Lévy process under the probability measure P on $(D, \mathscr{D})$. For all $t > 0$ we will denote P_t for the restriction of P to $\mathscr{D}_t$. In the case where $\int_{|y|\leq 1}|y|\nu(dy)<\infty$, we set $\gamma^\nu := \int_{|y|\leq 1} y\nu(dy)$. Note that, if ν is a finite Lévy measure, then the process having characteristic triplet $(\gamma^\nu, 0, \nu)$ is a compound Poisson process.

Here and in the sequel we will denote by Δx_r the jump of process $\{x_t\}$ at the time r:

$$\Delta x_r = x_r - \lim_{s\uparrow r} x_s.$$

For the proof of Theorems 4.2.5, 4.2.6 we also need some results on the equivalence of measures for Lévy processes. By the notation $\ll$ we will mean "is absolutely continuous with respect to".

Theorem 4.5.17 (See Sato (1999), Theorems 33.1–33.2 and Sato (2000) Corollary 3.18, Remark 3.19)**.** *Let P^1 (resp. P^2) be the law induced on $(D, \mathscr{D})$ by a Lévy process of characteristic triplet $(\eta, 0, \nu_1)$ (resp. $(0, 0, \nu_2)$), where*

$$\eta = \int_{|y|\leq 1} y(\nu_1 - \nu_2)(dy) \tag{4.26}$$

is supposed to be finite. Then $P_t^1 \ll P_t^2$ for all $t \geq 0$ if and only if $\nu_1 \ll \nu_2$ and the density $\frac{d\nu_1}{d\nu_2}$ satisfies

$$\int \left(\sqrt{\frac{d\nu_1}{d\nu_2}(y)} - 1\right)^2 \nu_2(dy) < \infty. \tag{4.27}$$

Remark that the finiteness in (4.27) *implies that in* (4.26)*. When $P_t^1 \ll P_t^2$, the density is*

$$\frac{dP_t^1}{dP_t^2}(x) = \exp(U_t(x)),$$

with

$$U_t(x) = \lim_{\varepsilon\to 0}\left(\sum_{r\leq t} \ln \frac{d\nu_1}{d\nu_2}(\Delta x_r)\mathbb{I}_{|\Delta x_r|>\varepsilon} - \int_{|y|>\varepsilon} t\left(\frac{d\nu_1}{d\nu_2}(y) - 1\right)\nu_2(dy)\right), P^{(0,0,\nu_2)}\text{-a.s.} \tag{4.28}$$

The convergence in (4.28) *is uniform in t on any bounded interval, $P^{(0,0,\nu_2)}$-a.s. Besides, $\{U_t(x)\}$ defined by* (4.28) *is a Lévy process satisfying $\mathbb{E}_{P^{(0,0,\nu_2)}}[e^{U_t(x)}] = 1$, $\forall t \geq 0$.*

Finally, let us consider the following result giving an explicit bound for the L_1 and the Hellinger distances between two Lévy processes of characteristic triplets of the form $(b_i, 0, \nu_i)$, $i = 1, 2$ with $b_1 - b_2 = \int_{|y|\leq 1} y(\nu_1 - \nu_2)(dy)$.

Theorem 4.5.18 (See Jacod, Shiryaev (2003))**.** *For any $0 < T < \infty$, let P_T^i be the probability measure induced on $(D, \mathscr{D}_T)$ by a Lévy process of characteristic triplet $(b_i, 0, \nu_i)$, $i = 1, 2$ and suppose that $\nu_1 \ll \nu_2$.*

If $H^2(\nu_1,\nu_2) := \int \Big(\sqrt{\frac{d\nu_1}{d\nu_2}(y)} - 1\Big)^2 \nu_2(dy) < \infty$, *then*

$$H^2(P_T^1, P_T^2) \leq \frac{T}{2} H^2(\nu_1,\nu_2).$$

We conclude the Appendix with a technical statement about the Le Cam distance for finite variation models.

Lemma 4.5.19.

$$\Delta(\mathscr{P}_n^{\nu_0}, \mathscr{P}_{n,FV}^{\nu_0}) = 0.$$

Proof. Consider the Markov kernels π_1, π_2 defined as follows

$$\pi_1(x,A) = \mathbb{I}_A(x^d), \quad \pi_2(x,A) = \mathbb{I}_A(x - \cdot\gamma^{\nu_0}), \quad \forall x \in D, A \in \mathscr{D},$$

where we have denoted by x^d the discontinuous part of the trajectory x, i.e. $\Delta x_r = x_r - \lim_{s\uparrow r} x_s$, $x_t^d = \sum_{r\leq t} \Delta x_r$ and by $x - \cdot\gamma^{\nu_0}$ the trajectory $x_t - t\gamma^{\nu_0}$, $t \in [0, T_n]$. On the one hand we have:

$$\begin{aligned}\pi_1 P^{(\gamma^{\nu-\nu_0},0,\nu)}(A) &= \int_D \pi_1(x,A) P^{(\gamma^{\nu-\nu_0},0,\nu)}(dx) = \int_D \mathbb{I}_A(x^d) P^{(\gamma^{\nu-\nu_0},0,\nu)}(dx)\\ &= P^{(\gamma^{\nu},0,\nu)}(A),\end{aligned}$$

where in the last equality we have used the fact that, under $P^{(\gamma^{\nu-\nu_0},0,\nu)}$, $\{x_t^d\}$ is a Lévy process with characteristic triplet $(\gamma^\nu, 0, \nu)$ (see Sato (1999), Theorem 19.3). On the other hand:

$$\begin{aligned}\pi_2 P^{(\gamma^{\nu},0,\nu)}(A) &= \int_D \pi_2(x,A) P^{(\gamma^{\nu_0},0,\nu)}(dx) = \int_D \mathbb{I}_A(x - \cdot\gamma^{\nu_0}) P^{(\gamma^{\nu},0,\nu)}(dx)\\ &= P^{(\gamma^{\nu-\nu_0},0,\nu)}(A),\end{aligned}$$

since, by definition, $\gamma^\nu - \gamma^{\nu_0}$ is equal to $\gamma^{\nu-\nu_0}$. The conclusion follows by the definition of the Le Cam distance. $\square$

Chapter 5

Asymptotic equivalence for density estimation and Gaussian white noise: an extension

Résumé Nous présentons au sein du Chapitre 5 une extension du résultat bien connu sur l'équivalence asymptotique entre un modèle à densité et un modèle de bruit blanc gaussien. Notre extension consiste à élargir la classe non paramétrique des densités possibles. Plus précisément, nous pouvons considérer des densités définies sur n'importe quel sous-intervalle de $\mathbb{R}$ aussi bien que des densités discontinues ou non bornées. Les résultats sont constructifs: toutes les équivalences asymptotiques sont établies en construisant des noyaux de Markov. Le Chapitre 5 est basé sur le preprint Mariucci 2015a.

Mot clés: Expériences statistiques non paramétriques, distance de Le Cam, estimation non paramétrique d'une densité, modèle de bruit blanc gaussien.

Abstract The aim of this chapter is to present an extension of the well-known asymptotic equivalence between density estimation experiments and a Gaussian white noise model. Our extension consists in enlarging the nonparametric class of the admissible densities. More precisely, we propose a way to allow densities defined on any subinterval of $\mathbb{R}$, and also some discontinuous or unbounded densities are considered (so long as the discontinuity and unboundedness patterns are somehow known a priori). The results are constructive: all the asymptotic equivalences are established by constructing explicit Markov kernels. This chapter is based on a submitted paper.

Keywords: Nonparametric experiments, Le Cam distance, nonparametric density estimation, Gaussian white noise.

5.1 Introduction

When looking for asymptotic results for some statistical model it is often useful to profit from a global asymptotic equivalence, in the Le Cam sense, in order to be allowed to work in a simpler but equivalent model. Indeed, proving an asymptotic equivalence result means that one can transfer asymptotic risk bounds for any inference problem from one model to the other, at least for bounded loss functions. Roughly speaking, saying that two models, $\mathscr{P}_1$ and $\mathscr{P}_2$, are equivalent means that they contain the same amount of information about the parameter that we are interested in. For the basic concepts and a detailed description of the notion of asymptotic equivalence, we refer to Le Cam (1986); Le Cam, Yang (2000). A short review of this topic will be given in the Appendix.

In recent years, numerous papers have been published on the subject of nonparametric asymptotic equivalence. For a non exhaustive list of the main ones among them, see, for example, the introduction in Mariucci (2015d). In this paper, we will focus on nonparametric density estimation experiments.

The seminal paper in this subject is due to Nussbaum (1996). There, the asymptotic equivalence between an experiment given by n observations of a density f on $[0,1]$ and a Gaussian white noise model:

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{n}}dW_t, \quad t \in [0,1],$$

was established. Over the years several generalizations of this result have been proposed such as Brown et al. (2004a); Carter (2002); Jähnisch, Nussbaum (2003). In Brown et al. (2004a), the authors obtained the global asymptotic equivalence between a Poisson process with variable intensity and a Gaussian white noise experiment with drift problem. Via Poissonization, this result was also extended to density estimation models. In Jähnisch, Nussbaum (2003) the authors proved the global asymptotic equivalence between a nonparametric model associated with the observation of independent but not identically distributed random variables on the unit interval and a bivariate Gaussian white noise model. More closely related to our work is the result in Carter (2002). In that paper, he proposed a new approach to establish the same normal approximations to density estimations experiments as in Nussbaum (1996). While the result in Nussbaum (1996) is obtained by means of Poissonization, in Carter (2002) the key step is to connect the density estimation problem to a multinomial experiment and to simplify the latter with a multivariate normal experiment.

The purpose of the present work is to generalize Nussbaum (1996) and Carter (2002). More precisely, the density estimation experiments that we consider consist of n independent observations $(Y_i)_{i=1}^n$ defined on a interval $I \subseteq \mathbb{R}$ from some unknown distribution P_f^g having density (with respect to the Lebesgue measure on I) $\frac{dP_f^g}{dx}(x) = f(x)g(x)$. In particular, we do not require $I \subseteq \mathbb{R}$ to be bounded as is generally done in the existing literature. The function g is supposed to be known whereas f is unknown and belongs to a certain nonparametric functional class $\mathscr{F}$. Formally, the statistical model we consider is

$$\mathscr{P}_n^g = \big(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \{\otimes_{i=1}^n P_f^g : f \in \mathscr{F}\}\big). \tag{5.1}$$

The exact assumptions on f and g will be specified in Section 5.2. Here, let us only stress the fact that f has to be bounded away from zero and infinity and sufficiently regular, whereas g can be both unbounded and discontinuous. The advantage with respect to the earlier works is that this framework allows us to treat densities of the form $h = fg$ not necessarily bounded nor smooth. See Section 5.3.1 for a discussion about the hypotheses.

Finally, let us introduce the Gaussian white noise model. For that, let us denote by $(C, \mathscr{C})$ the space of continuous mappings from I into $\mathbb{R}$ endowed with its standard filtration and by $\mathbb{W}_f^g$ the law induced on $(C, \mathscr{C})$ by a stochastic process satisfying:

$$dY_t = \sqrt{f(t)g(t)}dt + \frac{dW_t}{2\sqrt{n}}, \quad t \in I, \tag{5.2}$$

where $(W_t)_{t\in\mathbb{R}}$ is a Brownian motion on $\mathbb{R}$ conditional on $W_0 = 0$. Then we set

$$\mathscr{W}_n^g = \big(C, \mathscr{C}, \{\mathbb{W}_f^g : f \in \mathscr{F}\}\big). \tag{5.3}$$

Let Δ be the Le Cam pseudo-distance between statistical models having the same parameter space. For the convenience of the reader a formal definition is given in Section 5.4.2. Our main result is then as follows (see Theorem 5.3.1 for the precise statement):

Main result 5.1.1. *Let I be a possibly infinite subinterval of $\mathbb{R}$ and let $\mathscr{F}$ consist of functions bounded away from 0 and ∞, satisfying the regularity assumptions stated in Section 5.2. Then, we have*

$$\lim_{n\to\infty} \Delta(\mathscr{P}_n^g, \mathscr{W}_n^g) = 0. \tag{5.4}$$

In some special cases an explicit upper bound for the rate of convergence in (5.4) is available; see, e.g. Corollary 5.3.2. The structure of the proof follows Carter's in

Carter (2002), but we detach from it on several aspects. The basic idea is to use his multinomial-multivariate normal approximation, but some technical points have to be taken into account. One of these is that I may be infinite, so that, in particular, the subintervals J_i in which it is partitioned cannot be of equal length. We choose intervals J_i of varying length, according to the quantiles of ν_0, the measure having density g with respect to Lebesgue. This kind of partitions was already considered in Mariucci (2015d).

The paper is organized as follows. Section 5.2 fixes the assumptions on the parameter space $\mathscr{F}$. Section 5.3 contains the statement of the main results and a discussion while Section 5.4 is devoted to the proofs. The paper includes an Appendix recalling the definition and some useful properties of the Le Cam distance.

5.2 The parameter space

Fix a finite measure ν_0 concentrated on a possibly infinite interval $I \subset \mathbb{R}$, admitting a density $g > 0$ with respect to Lebesgue. The class of functions $\mathscr{F}$ will be considered as a class of probability densities with respect to ν_0, i.e. $\int_I f(x)g(x)dx = 1$. For each $f \in \mathscr{F}$, let ν (resp. $\hat{\nu}_m$) be the measure having f (resp. $\hat{f}_m$) as a density with respect to ν_0 where, for every $f \in \mathscr{F}$, $\hat{f}_m(x)$ is defined as follows. Given a positive integer m, let $J_1 = I \cap (-\infty, v_1]$, $J_j := (v_{j-1}, v_j]$ for $j = 2, \dots, m-1$ and $J_m = I \cap (v_m, \infty)$ where the v_j's are the quantiles for ν_0, i.e.

$$\mu_n := \nu_0(J_j) = \frac{\nu_0(I)}{m}, \quad \forall j = 1, \dots, m. \tag{5.5}$$

Define $x_j^* := \frac{\int_{J_j} x\nu_0(dx)}{\mu_n}$, $j = 1, \dots, m$ and introduce a sequence of continuous functions $0 \le V_j \le \frac{1}{\mu_n}$, $j = 1, \dots, m$, defined in the following way.

- V_1 is supported on $I \cap (-\infty, x_2^*]$ and:

$$V_1|_{I\cap(-\infty,x_1^*]} \equiv \frac{1}{\mu_n}; \quad \int_{x_1^*}^{x_2^*} V_1(x)\nu_0(dx) = \frac{\nu_0((x_1^*, v_1])}{\mu_n}; \quad V_1(x_2^*) = 0.$$

- For $j = 2, \dots, m-1$, V_j is supported in $[x_{j-1}^*, x_{j+1}^*]$ and:

$$V_j|_{[x_{j-1}^*, x_j^*]} \equiv 1 - V_{j-1}|_{[x_{j-1}^*, x_{j+1}^*]}; \quad \int_{x_j^*}^{x_{j+1}^*} V_j(x)\nu_0(dx) = \frac{\nu_0((x_j^*, v_j])}{\mu_n}; \quad V_j(x_{j+1}^*) = 0.$$

- For $j = m$, V_m is supported on $I \setminus (-\infty, x_{m-1}^*)$ and:

$$V_m|_{[x_{m-1}^*, x_m^*]} \equiv 1 - V_{m-1}|_{[x_{m-1}^*, x_m^*]} \quad \text{and} \quad V|_{I\cap(-\infty,x_m^*)} \equiv \frac{1}{\mu_n}.$$

We now explain the assumptions we will need to make on the parameter f. We require that:

(H1) There exist constants $\kappa, M > 0$ such that $\kappa \leq f(y) \leq M$, for all $y \in I$ and $f \in \mathscr{F}$.

The m introduced above will be considered as a function of n, $m = m_n$. We can thus consider $\widehat{\sqrt{f}}_m$, the approximation of $\sqrt{f}$ constructed as $\hat{f}_m$ above and introduce the quantities:

$$\begin{aligned}
H_m^2(f) &:= \int_I \Big(\sqrt{f(x)} - \sqrt{\hat{f}_m(x)}\Big)^2 \nu_0(dx),\\
A_m^2(f) &:= \int_I \Big(\widehat{\sqrt{f}}_m(y) - \sqrt{f(y)}\Big)^2 \nu_0(dy),\\
B_m^2(f) &:= \sum_{j=1}^m \bigg(\int_{J_j} \frac{\sqrt{f(y)}}{\sqrt{\nu_0(J_j)}} \nu_0(dy) - \sqrt{\nu(J_j)}\bigg)^2.
\end{aligned}$$

We will assume the existence of a sequence of discretizations $m = m_n$ and functions V_j, $j = 1, \dots, m$, such that:

(C1) $\lim_{n\to\infty} n \sup_{f\in\mathscr{F}} \big(H_m^2(f) + A_m^2(f) + B_m^2(f)\big) = 0.$

5.3 Main results and discussion

Using the notation introduced in Section 5.2, we now state our main result in terms of the models $\mathscr{P}_n^g$ and $\mathscr{W}_n^g$ defined in (5.1) and (5.3), respectively.

Theorem 5.3.1. *Let ν_0 be a finite measure concentrated on an (possibly infinite) interval $I \subset \mathbb{R}$ having density $g > 0$ with respect to Lebesgue. Suppose that there exist a sequence $m = m_n$ and functions V_j, $j = 1, \dots, m$, such that every $f \in \mathscr{F}$ satisfies conditions (H1) and (C1). Then, for n big enough we have:*

$$\Delta(\mathscr{P}_n^g, \mathscr{W}_n^g) = O\bigg(\sqrt{n} \sup_{f\in\mathscr{F}} \Big(A_m(f) + B_m(f) + H_m(f)\Big) + \frac{m \ln m}{\sqrt{n}}\bigg).$$

Corollary 5.3.2. *Let I be a compact subset of $\mathbb{R}$ and let ν_0 be the Lebesgue measure on I. For fixed $\gamma \in (0, 1]$ and K, κ, M strictly positive constants, consider the functional class*

$$\mathscr{F}_{(\gamma,K,\kappa,M)} = \Big\{ f \in C^1(I) : \kappa \leq f(x) \leq M,\ |f'(x) - f'(y)| \leq K|x-y|^\gamma,\ \forall x, y \in I \Big\}.$$

Suppose $\mathscr{F} \subset \mathscr{F}_{(\gamma,K,\kappa,M)}$. Then

$$\Delta(\mathscr{P}_n^g, \mathscr{W}_n^g) = O\Big(n^{-\frac{\gamma}{\gamma+2}} \log n\Big).$$

5.3.1 Existing literature and discussion

As it has already been highlighted in the introduction, our result is a generalization of those in Nussbaum (1996) and Carter (2002). In order to discuss the link between our work and the previous ones, we recall the results contained in these papers.

- *Asymptotic equivalence of density estimation and Gaussian white noise*, Nussbaum (1996).

  In this paper Nussbaum establishes a global asymptotic equivalence between the problem of density estimation from an i.i.d. sample and a Gaussian white noise model. More precisely, let $(Y_i)_{i=1}^n$ be i.i.d. random variables with density f on $[0,1]$ with respect to the Lebesgue measure. The densities f are the unknown parameters and they are supposed to belong to a certain nonparametric class $\mathscr{F}$ subject to a Hölder restriction: $|f(x)-f(y)| \leq C|x-y|^\alpha$ with $\alpha > \frac{1}{2}$ and a positivity restriction: $f(x) \geq \varepsilon > 0$. Let us denote by $\mathscr{P}_{1,n}$ the statistical model associated with the observation of the Y_i's. Furthermore, let $\mathscr{P}_{2,n}$ be the experiment in which one observes a stochastic process $(Y_t)_{t\in[0,1]}$ such that

  $$dY_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{n}}dW_t, \quad t \in [0,1],$$

  where $(W_t)_{t\in[0,1]}$ is a standard Brownian motion. Then the main result in Nussbaum (1996) is that $\Delta(\mathscr{P}_{1,n}, \mathscr{P}_{2,n}) \to 0$ as $n \to \infty$.

  This is done by first showing that the result holds for certain subsets $\mathscr{F}_n(f_0)$ of the class $\mathscr{F}$ described above. Then it is shown that one can estimate the f_0 rapidly enough to fit the various pieces together. Without entering into any detail, let us just mention that the key steps are a Poissonization technique and the use of a functional KMT inequality.

- *Deficiency distance between multinomial and multivariate normal experiments,* Carter (2002).

  In this paper Carter establishes a global asymptotic equivalence between a density estimation model and a Gaussian white noise model by bounding the Le Cam distance between multinomial and multivariate normal random variables. More precisely, let us denote by $\mathcal{M}(n,\theta)$ the multinomial distribution, where $\theta := (\theta_1,\dots,\theta_m)$. Denote the covariance matrix nV_θ: Its (i,j)th element equals to $n\theta_i(1-\theta_i)\delta_{i,j}-n\theta_i\theta_j$.

  The main result is an upper bound for the Le Cam distance $\Delta(\mathcal{M},\mathcal{N})$ between the models $\mathcal{M} := \{\mathcal{M}(n,\theta) : \theta \in \Theta\}$ and $\mathcal{N} := \{\mathscr{N}(n\theta, nV_\theta) : \theta \in \Theta\}$, under some regularity assumptions on Θ. In particular, Carter proves that

$$\Delta(\mathcal{M},\mathcal{N}) \leq C'_\Theta \frac{m \ln m}{\sqrt{n}} \quad \text{provided} \quad \sup_{\theta\in\Theta} \frac{\max_i \theta_i}{\min_i \theta_i} \leq C_\Theta < \infty,$$

for a constant C'_Θ that depends only on C_Θ. From this inequality Carter can recover mostly the same results as Nussbaum (1996) under stronger regularity assumptions on $\mathscr{F}$: $\mathscr{F}$ is a class of smooth, differentiable densities f on the interval $[0,1]$ such that there exist strictly positive constants ε, M, γ such that $\varepsilon \leq f \leq M$ and

$$|f'(x) - f'(y)| \leq M|x-y|^\gamma, \quad \text{for all } x,y \in [0,1].$$

Let us briefly explain how one can use a bound on the distance between multinomial and multivariate normal variables to make assertions about density estimation experiments. The idea is to see the multinomial experiment as the result of grouping independent observations from a continuous density into subsets. Using the square root as a variance-stabilizing transformation, these multinomial variables can be asymptotically approximated by normal variables with constant variances. These normal variables, in turn, are approximations to the increments of the Brownian motion processes over the sets in the partition.

Our work can be seen as a generalization of the previously cited works: To see that it is enough to take $g(x) = \mathbb{I}_{[0,1]}(x)$ as in Corollary 5.3.2. However, it differs from Nussbaum and Carter's results in several aspects. First of all, we do not need to ask the random variables to be defined on $[0,1]$, allowing the observations to be defined on a possibly infinite interval I of $\mathbb{R}$. Secondly, in our setting the positivity restriction on the densities can be removed. Indeed, as a parametric example, we can consider truncated Gamma distributions on $[0,L]$, that is distributions having a density h with respect to the Lebesgue measure:

$$h(x) = \frac{\exp(-\theta x)\theta^n x^{n-1}}{\int_0^L \exp(-\theta y)\theta^n y^{n-1} dy} \mathbb{I}_{[0,L]}(x).$$

We can apply Theorem 5.3.1, taking $\mathscr{F} = \{f_\theta : \theta \in \mathbb{R}_{>0}\}$ and

$$f_\theta(x) = \frac{\exp(-\theta x)\theta^n}{\int_0^L \exp(-\theta y)\theta^n y^{n-1} dy} \mathbb{I}_{[0,L]}(x), \quad g(x) = x^{n-1}.$$

More generally, density functions h that can be written in form of a product are commonly used in statistics. Again, one could cite as a simple case the problem of a parametric estimation for a Weibull density, see, e.g. Diebolt et al. (2008); Gardes, Girard (2006). Generally speaking, the present work can be useful whenever the random variables Y_i's do not admit a smooth density h with respect to Lebesgue, but nevertheless one has some informations on the discontinuity structure, namely one knows g in the decomposition $h(x) = f(x)g(x)$.

5.4 Proofs

5.4.1 Proof of Theorem 5.3.1

We will proceed in four steps.

Step 1. By means of Facts 5.4.4 and 5.4.5, we get

$$\Big\|\bigotimes_{i=1}^n P_f^g - \bigotimes_{i=1}^n P_{\hat{f}_m}^g\Big\|_{TV} \leq H\bigg(\bigotimes_{i=1}^n P_f^g, \bigotimes_{i=1}^n P_{\hat{f}_m}^g\bigg) \leq \sqrt{nH^2\big(P_f^g, P_{\hat{f}_m}^g\big)}.$$

Hence, denoting by $\hat{\mathscr{P}}_n^g$ the statistical model associated with the family of probabilities $\big\{P_{\hat{f}_m}^g; f \in \mathscr{F}\big\}$:

$$\Delta(\mathscr{P}_n^g, \hat{\mathscr{P}}_n^g) \leq \sup_{f\in\mathscr{F}} \sqrt{n\int_I \bigg(\sqrt{f(x)} - \sqrt{\hat{f}_m(x)}\bigg)^2 g(x)dx}. \tag{5.6}$$

Step 2. Following the same approach as in Carter (2002), we introduce an auxiliary multinomial experiment to get closer to a normal one representing the increments of $(Y_t)_{t\in I}$ defined as in (5.2). The multinomial experiment is linked with the density estimation model in the following way: Let $(Y_i)_{i=1}^n$ be a sequence of i.i.d. random variables with density fg with respect to Lebesgue and define the multinomial experiment by grouping their observations into subsets. More precisely, let us introduce the random variables:

$$Z_i = \sum_{j=1}^n \mathbb{I}_{J_i}(Y_j),\ \ i = 1, \dots, m.$$

Observe that the law of the vector $(Z_1, \dots, Z_m)$ is multinomial $\mathcal{M}(n; \gamma_1, \dots, \gamma_m)$ where

$$\gamma_i = \int_{J_i} f(x)g(x)dx, \quad i = 1, \dots, m.$$

Let us denote by $\mathcal{M}_m$ the statistical model associated with the observation of $(Z_1, \dots, Z_m)$. Clearly $\delta(\mathscr{P}_n^g, \mathcal{M}_m) = 0$. Indeed, $\mathcal{M}_m$ is the image experiment by the random variable $S : I^n \to \{1, \dots, n\}^m$ defined as

$$S(x_1, \dots, x_n) = \Big(\#\big\{j : x_j \in J_1\big\}; \dots; \#\big\{j : x_j \in J_m\big\}\Big),$$

where $\#A$ denotes the cardinal of the set A. To conclude the second step we now prove that the multinomial experiment is as informative as $\hat{\mathscr{P}}_n^g$.

Lemma 5.4.1.

$$\delta(\mathcal{M}_m, \hat{\mathscr{P}}_n^g) = 0.$$

Proof. We need to produce an explicit Markov kernel that allows to approximate the density $\hat{f}_m g$ given an observation from the multinomial model which is given by

$$K\big((k_1,\dots,k_m),A\big) = \int_A \mathbb{E}\Big[V_{X_{(k_1,\dots,k_m)}}(x)\Big]\nu_0(dx), \quad \forall (k_1,\dots,k_m)\in\mathbb{N},\ \sum_i k_i = n,\ A\subset\mathbb{R},$$

where $X_{(k_1,\dots,k_m)} \in \{1,\dots,m\}$ is a randomly chosen integer obtained assigning to j the weight $\frac{k_j}{n}$. □

Step 3. Let us denote by $\mathcal{N}_m$ the statistical model associated with the observation of m independent Gaussian variables $\mathscr{N}(\sqrt{n\gamma_i}, \frac{1}{4})$, $i = 1,\dots,m$. Since $\frac{\max \gamma_i}{\min \gamma_i} \leq \frac{M}{\kappa}$, one can apply Theorem 5.4.9 obtaining

$$\Delta(\mathcal{M}_m, \mathcal{N}_m) = O\Big(\frac{m \ln m}{\sqrt{n}}\Big).$$

Here the O depends only on M and κ.

Step 4. Finally, we conclude the proof of Theorem 5.3.1, by showing that

$$\Delta(\mathscr{N}_m, \mathscr{W}_n^g) \leq 2\sqrt{n} \sup_{f\in\mathscr{F}} \big(A_m(f) + B_m(f)\big). \tag{5.7}$$

As a preliminary remark note that $\mathscr{W}_n^g$ is equivalent to the model that observes a trajectory from:

$$d\bar{y}_t = \sqrt{f(t)}g(t)dt + \frac{\sqrt{g(t)}}{2\sqrt{n}}dW_t, \quad t\in I.$$

In order to prove (5.7) we proceed in the following way: First of all, we prove that $\mathscr{N}_m$ is equivalent to the model that observes the increments on the intervals J_i of $(\bar{y}_t)_{t\in I}$. Secondly, we show that the increments of $(\bar{y}_t)_{t\in I}$ are more informative than another Gaussian process, say $(Y_t^*)_{t\in I}$, that turns out to be very close to $(\bar{y}_t)_{t\in I}$ in the total variation distance. We then conclude the asymptotic equivalence between $\mathscr{N}_m$ and $\mathscr{W}_n^g$ observing that the increments of $(\bar{y}_t)_{t\in I}$ are obviously less informative than $\mathscr{W}_n^g$.

Let us denote by $\bar{Y}_j$ the increments of the process $(\bar{y}_t)$ over the intervals J_j, $j = 1,\dots,m$, i.e.

$$\bar{Y}_j := \bar{y}_{v_j} - \bar{y}_{v_{j-1}} \sim \mathscr{N}\bigg(\int_{J_j}\sqrt{f(y)}\nu_0(dy), \frac{\nu_0(J_j)}{4n}\bigg)$$

and denote by $\bar{\mathscr{N}}_m$ the statistical model associated with the distributions of these increments. As announced we start by bounding the Le Cam distance between $\mathscr{N}_m$ and $\bar{\mathscr{N}}_m$ showing that

$$\Delta(\mathscr{N}_m,\bar{\mathscr{N}}_m)\leq 2\sqrt{n}\sup_{f\in\mathscr{F}}B_m(f),\quad\text{for all m.}\tag{5.8}$$

In this regard, remark that the experiment $\bar{\mathscr{N}}_m$ is equivalent to another experiment, say $\mathscr{N}_m^{\#}$, that observes m independent Gaussian random variables of means equal to $\frac{2\sqrt{n}}{\sqrt{\nu_0(J_j)}}\int_{J_j}\sqrt{f(y)}\nu_0(dy)$, $j=1,\dots,m$ and variances identically 1. Hence, using also Property 5.4.3, Facts 5.4.4 and 5.4.6 we get:

$$\Delta(\mathscr{N}_m,\bar{\mathscr{N}}_m)\leq\Delta(\mathscr{N}_m,\mathscr{N}_m^{\#})\leq\sqrt{\sum_{j=1}^m\bigg(\frac{2\sqrt{n}}{\sqrt{\nu_0(J_j)}}\int_{J_j}\sqrt{f(y)}\nu_0(dy)-2\sqrt{n\nu(J_j)}\bigg)^2}.$$

Using similar ideas as in Section 8.2 of Carter (2002) and Lemma 3.2 of Mariucci (2015d), we introduce a new stochastic process constructed from the random variables $\bar{Y}_j$'s. To that end define

$$Y_t^*=\sum_{j=1}^m\bar{Y}_j\int_{I\cap(-\infty,t]}V_j(y)\nu_0(dy)+\frac{1}{2\sqrt{n}}\sum_{j=1}^m\sqrt{\nu_0(J_j)}B_j(t),\quad t\in I,\tag{5.9}$$

where the $(B_j(t))_t$ are independent centered Gaussian processes conditional on $B_j(0)=0$ with variances

$$\mathrm{Var}(B_j(t))=\int_{I\cap(-\infty,t]}V_j(y)\nu_0(dy)-\bigg(\int_{I\cap(-\infty,t]}V_j(y)\nu_0(dy)\bigg)^2.$$

By construction, (Y_t^*) is a Gaussian process with mean and variance given by, respectively:

$$\begin{aligned}\mathbb{E}[Y_t^*]&=\sum_{j=1}^m\mathbb{E}[\bar{Y}_j]\int_{I\cap(-\infty,t]}V_j(y)\nu_0(dy)=\sum_{j=1}^m\bigg(\int_{J_j}\sqrt{f(y)}\nu_0(dy)\bigg)\int_{I\cap(-\infty,t]}V_j(y)\nu_0(dy),\\ \mathrm{Var}[Y_t^*]&=\sum_{j=1}^m\mathrm{Var}[\bar{Y}_j]\bigg(\int_{I\cap(-\infty,t]}V_j(y)\nu_0(dy)\bigg)^2+\frac{1}{4n}\sum_{j=1}^m\nu_0(J_j)\mathrm{Var}(B_j(t))\\ &=\frac{1}{4n}\int_{I\cap(-\infty,t]}\sum_{j=1}^m\nu_0(J_j)V_j(y)\nu_0(dy)=\frac{1}{4n}\int_{I\cap(-\infty,t]}1\nu_0(dy)=\frac{\nu_0(I\cap(-\infty,t])}{4n}.\end{aligned}$$

Therefore,

$$Y_t^*=\int_{I\cap(-\infty,t]}\widehat{\sqrt{f}}_m(y)\nu_0(dy)+\int_{I\cap(-\infty,t]}\frac{\sqrt{g(t)}}{2\sqrt{n}}W_t,\quad t\in I,$$

where

$$\widehat{\sqrt{f}}_m(x) := \sum_{j=1}^{m} \Big(\int_{J_j} \sqrt{f(y)} \nu_0(dy) \Big) V_j(x).$$

Applying Fact 5.4.7, we get that the total variation distance between the process $(Y_t^*)_{t\in I}$ constructed from the random variables $\bar{Y}_j$, $j = 1, \dots, m$ and the Gaussian process $(\bar{y}_t)_{t\in I}$ is bounded by

$$\sqrt{4n \int_I \big(\widehat{\sqrt{f}}_m(y) - \sqrt{f(y)}\big)^2 \nu_0(dy)},$$

as wanted.

5.4.2 Proof of Corollary 5.3.2

We start by proving a Lemma needed for the proof of Corollary 5.3.2. Since we are supposing that $g(x) = \mathbb{I}_I(x)$, we may take for the V_j the standard choice of triangular-trapezoidal functions (see Example 4.2.1 for a picture). Furthermore, $\mu_n = \nu_0(J_j) = \frac{1}{m}|I|$. For the easiness of notations, in the proof we will also assume $I = [0, 1]$.

Lemma 5.4.2. *If $f \in \mathscr{F}_{(\gamma,K,\kappa,M)}$ then*

$$\|f - \hat{f}_m\|^2_{L_2(\nu_0)} \leq O\Big(m^{-3} + m^{-2-2\gamma}\Big),$$

with the O depending on K, M and κ.

Proof. Let us consider the Taylor expansion of f at points x_j^*, where x denotes a point in $(x_{j-1}^*, x_j^*]$, $j = 2, \dots, m$:

$$f(x) = f(x_j^*) + f'(x_j^*)(x - x_j^*) + R(x). \tag{5.10}$$

The smoothness condition on f allows us to bound the error R as follows:

$$\begin{aligned} |R(x)| &= \Big|f(x) - f(x_j^*) - f'(x_j^*)(x - x_j^*)\Big| \\ &= \big|f'(\xi_j) - f'(x_j^*)\big||\xi_j - x_j^*| \leq K m^{-1-\gamma}, \end{aligned}$$

where ξ_j is a certain point in $(x_{j-1}^*, x_j^*]$.

By the linear character of $\hat{f}_m$, we can write:

$$\hat{f}_m(x) = \hat{f}_m(x_j^*) + \hat{f}'_m(x_j^*)(x - x_j^*)$$

where $\hat{f}'_m$ denotes the left or right derivative of $\hat{f}_m$ in x_j^* depending whether $x < x_j^*$ or $x > x_j^*$; this equals to $f'(t)$ for some $t \in J_j \cup J_{j+1}$, which allows us to exploit the Hölder condition. Indeed, if $x \in J_j$, $j = 1, \dots, m$, then there exists $t \in J_j \cup J_{j+1}$ such that:

$$\begin{aligned}|f(x) - \hat{f}_m(x)| &\leq |f(x_j^*) - \hat{f}_m(x_j^*)| + |f'(x_j^*) - f'(t)||t - x_j^*| + |R(x)| \\ &\leq |f(x_j^*) - \hat{f}_m(x_j^*)| + K|t - x_j^*|^{\gamma+1} + Km^{-1-\gamma} \leq |f(x_j^*) - \hat{f}_m(x_j^*)| + 3Km^{-1-\gamma}.\end{aligned}$$

Using (5.10) and the fact that $\int_{J_j}(x - x_j^*)\nu_0(dx) = 0$, one gets:

$$\big|f(x_j^*) - \hat{f}_m(x_j^*)\big| = \frac{1}{\nu_0(J_j)}\bigg|\int_{J_j}\big(f(x_j^*) - f(x)\big)\nu_0(dx)\bigg| \leq Km^{-1-\gamma}.$$

Moreover, observe that, for all $x \in J_i$, $i = 1, \dots, m$, $\big|f(x) - \frac{\nu(J_j)}{\nu_0(J_j)}\big|$, is bounded by $3Km^{-1-\gamma} + m^{-1}M$, indeed:

$$\begin{aligned}\bigg|f(x) - \frac{\nu(J_j)}{\nu_0(J_j)}\bigg| &= |f(x) - \hat{f}_m(x_i^*)| \leq |f(x) - \hat{f}_m(x)| + |\hat{f}_m(x) - \hat{f}_m(x_i^*)| \\ &\leq 3Km^{-1-\gamma} + |\hat{f}'_m(x_i^*)(x - x_i^*)| \leq 3Km^{-1-\gamma} + Mm^{-1}.\end{aligned}$$

Collecting all the pieces together we find

$$\int_I \big(f(x) - \hat{f}_m(x)\big)^2\nu_0(dx) \leq 2m^{-1}n\Big(3Km^{-1-\gamma} + Mm^{-1}\Big)^2 + 18K^2m^{-2-2\gamma}.$$

□

Proof of Corollary 5.3.2. By means of the fact that $f(x) \geq \kappa$ for all $x \in I$ one can write:

$$\begin{aligned}\int_I\Big(\sqrt{f(x)} - \sqrt{\hat{f}_m(x)}\Big)^2 dx &= \int_I\bigg(\frac{f(x) - \hat{f}_m(x)}{\sqrt{f(x)} + \sqrt{\hat{f}_m(x)}}\bigg)^2 dx \\ &\leq \frac{1}{4\kappa}\int_I\big(f(x) - \hat{f}_m(x)\big)^2 dx.\end{aligned}$$

A straightforward application of Lemma 5.4.2 gives

$$H_m^2(f) = O\Big(m^{-3} + m^{-2-2\gamma}\Big).$$

The same bound holds for $A_m^2(f)$ since if $f \in \mathscr{F}_{(\gamma,K,\kappa,M)}$ then $\sqrt{f} \in \mathscr{F}_{(\gamma,\frac{K}{\sqrt{\kappa}},\sqrt{\kappa},\sqrt{M})}$. Moreover, one can see that B_m converges with the same rate as A_m. This may be done by explicit computations, see Mariucci (2015d), Lemma 3.10 for more details. □

Background

Le Cam theory of statistical experiments

A *statistical model* or *experiment* is a triplet $\mathscr{P}_j = (\mathscr{X}_j, \mathscr{A}_j, \{P_{j,\theta}; \theta \in \Theta\})$ where $\{P_{j,\theta}; \theta \in \Theta\}$ is a family of probability distributions all defined on the same σ-field $\mathscr{A}_j$ over the *sample space* $\mathscr{X}_j$ and Θ is the *parameter space*. The *deficiency* $\delta(\mathscr{P}_1, \mathscr{P}_2)$ of $\mathscr{P}_1$ with respect to $\mathscr{P}_2$ quantifies "how much information we lose" by using $\mathscr{P}_1$ instead of $\mathscr{P}_2$ and it is defined as $\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_K \sup_{\theta\in\Theta} ||KP_{1,\theta} - P_{2,\theta}||_{TV}$, where TV stands for "total variation" and the infimum is taken over all "transitions" K (see Le Cam (1986), page 18). The general definition of transition is quite involved but, for our purposes, it is enough to know that (possibly randomized) Markov kernels are special cases of transitions. By $KP_{1,\theta}$ we mean the image measure of $P_{1,\theta}$ via the Markov kernel K, that is

$$KP_{1,\theta}(A) = \int_{\mathscr{X}_1} K(x, A)P_{1,\theta}(dx), \quad \forall A \in \mathscr{A}_2.$$

The experiment $K\mathscr{P}_1 = (\mathscr{X}_2, \mathscr{A}_2, \{KP_{1,\theta}; \theta \in \Theta\})$ is called a *randomization* of $\mathscr{P}_1$ by the Markov kernel K. When the kernel K is deterministic, that is $K(x, A) = \mathbb{I}_A S(x)$ for some random variable $S : (\mathscr{X}_1, \mathscr{A}_1) \to (\mathscr{X}_2, \mathscr{A}_2)$, the experiment $K\mathscr{P}_1$ is called the *image experiment by the random variable* S. The Le Cam distance is defined as the symetrization of δ and it defines a pseudometric. When $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$ the two statistical models are said to be *equivalent*. Two sequences of statistical models $(\mathscr{P}_1^n)_{n\in\mathbb{N}}$ and $(\mathscr{P}_2^n)_{n\in\mathbb{N}}$ are called *asymptotically equivalent* if $\Delta(\mathscr{P}_1^n, \mathscr{P}_2^n)$ tends to zero as n goes to infinity. A very interesting feature of the Δ-distance is that it can be also translated in terms of statistical decision theory. Let $\mathscr{D}$ be any (measurable) decision space and let $L : \Theta \times \mathscr{D} \mapsto [0, \infty)$ denote a loss function. Let $\|L\| = \sup_{(\theta,z)\in\Theta\times\mathscr{D}} L(\theta, z)$. Let π_i denote a (randomized) decision procedure in the i-th experiment. Denote by $R_i(\pi_i, L, \theta)$ the risk from using procedure π_i when L is the loss function and θ is the true value of the parameter. Then, an equivalent definition of the deficiency is:

$$\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_{\pi_1} \sup_{\pi_2} \sup_{\theta\in\Theta} \sup_{L:\|L\|=1} \big|R_1(\pi_1, L, \theta) - R_2(\pi_2, L, \theta)\big|.$$

Thus $\Delta(\mathscr{P}_1, \mathscr{P}_2) < \varepsilon$ means that for every procedure π_i in problem i there is a procedure π_j in problem j, $\{i, j\} = \{1, 2\}$, with risks differing by at most ε, uniformly over all bounded L and $\theta \in \Theta$. In particular, when minimax rates of convergence in a nonparametric estimation problem are obtained in one experiment, the same rates automatically hold in any asymptotically equivalent experiment. There is more: When explicit transformations from one experiment to another are obtained, statistical procedures can be carried over from one experiment to the other one.

There are various techniques to bound the Le Cam distance. We report below only the properties that are useful for our purposes. For the proofs see, e.g., Le Cam (1986); Strasser (1985).

Property 5.4.3. *Let $\mathscr{P}_j = (\mathscr{X}, \mathscr{A}, \{P_{j,\theta}; \theta \in \Theta\})$, $j = 1, 2$, be two statistical models having the same sample space and define $\Delta_0(\mathscr{P}_1, \mathscr{P}_2) := \sup_{\theta\in\Theta} \|P_{1,\theta} - P_{2,\theta}\|_{TV}$. Then, $\Delta(\mathscr{P}_1, \mathscr{P}_2) \leq \Delta_0(\mathscr{P}_1, \mathscr{P}_2)$.*

In particular, Property 5.4.3 allows us to bound the Le Cam distance between statistical models sharing the same sample space by means of classical bounds for the total variation distance. To that aim, we collect below some useful results.

Fact 5.4.4. *Let P_1 and P_2 be two probability measures on $\mathscr{X}$, dominated by a common measure ξ, with densities $g_i = \frac{dP_i}{d\xi}$, $i = 1, 2$. Define*

$$L_1(P_1, P_2) = \int_{\mathscr{X}} |g_1(x) - g_2(x)|\xi(dx),$$

$$H(P_1, P_2) = \bigg(\int_{\mathscr{X}} \Big(\sqrt{g_1(x)} - \sqrt{g_2(x)}\Big)^2 \xi(dx)\bigg)^{1/2}.$$

Then,

$$\frac{H^2(P_1, P_2)}{2} \leq \|P_1 - P_2\|_{TV} = \frac{1}{2}L_1(P_1, P_2) \leq H(P_1, P_2).$$

Fact 5.4.5. *Let P and Q be two product measures defined on the same sample space: $P = \otimes_{i=1}^n P_i$, $Q = \otimes_{i=1}^n Q_i$. Then*

$$H^2(P, Q) \leq \sum_{i=1}^n H^2(P_i, Q_i).$$

Fact 5.4.6. *Let $Q_1 \sim \mathscr{N}(\mu_1, \sigma_1^2)$ and $Q_2 \sim \mathscr{N}(\mu_2, \sigma_2^2)$. Then*

$$\|Q_1 - Q_2\|_{TV} \leq \sqrt{2\bigg(1 - \frac{\sigma_1^2}{\sigma_2^2}\bigg)^2 + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}}.$$

Fact 5.4.7. *For $i = 1, 2$, let Q_i, $i = 1, 2$, be the law on $(C, \mathscr{C})$ of two Gaussian processes of the form*

$$X_t^i = \int_0^t h_i(s)ds + \int_0^t \sigma(s)dW_s,\ t \in I$$

where $h_i \in L_2(\mathbb{R})$ and $\sigma \in \mathbb{R}_{>0}$. Then:

$$L_1\big(Q_1, Q_2\big) \leq \sqrt{\int_I \frac{\big(h_1(s) - h_2(s)\big)^2}{\sigma^2(s)}ds}.$$

Property 5.4.8. *Let $\mathscr{P}_i = (\mathscr{X}_i, \mathscr{A}_i, \{P_{i,\theta}, \theta \in \Theta\})$, $i = 1, 2$, be two statistical models. Let $S : \mathscr{X}_1 \to \mathscr{X}_2$ be a sufficient statistics such that the distribution of S under $P_{1,\theta}$ is equal to $P_{2,\theta}$. Then $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$.*

Finally, we recall the following result that allows us to bound the Le Cam distance between multinomial and Gaussian variables. According with the notation used throughout the paper, $\mathscr{M}(n, \theta)$ stands for a multinomial distribution of parameters (n, θ).

Theorem 5.4.9. *(See Carter (2002), Theorem 1 and Sections 7.1, 7.2) Let $\mathscr{P} = \{P_\theta : \theta \in \Theta_R\}$, where $P_\theta = \mathscr{M}(n, \theta)$ and $\Theta_R \subset \mathbb{R}^m$ consists of all vectors of probabilities such that*

$$\frac{\max \theta_i}{\min \theta_i} \leq R.$$

Let $\mathscr{Q} = \{Q_\theta : \theta \in \Theta_R\}$ where Q_θ is the multivariate normal distribution with vector mean $(\sqrt{n\theta_1}, \dots, \sqrt{n\theta_m})$ and diagonal covariance matrix $\frac{1}{4} I_m$. Then

$$\Delta(\mathscr{P}, \mathscr{Q}) \leq C_R \frac{m \ln m}{\sqrt{n}}$$

for a constant C_R that depends only on R.

Chapter 6

Asymptotic equivalence of discretely observed diffusion processes and their Euler scheme: small variance case

Résumé L'objet du Chapitre 6 est un résultat d'équivalence asymptotique entre un modèle de diffusion scalaire avec une dérive inconnue et un coefficient de diffusion qui tend vers zéro et le schéma d'Euler associé. L'horizon temporel $T < \infty$ est fixé et les deux cas d'observations discrètes ou continues sont étudiés. Le coefficient de diffusion peut être non constant. Toutes les équivalences asymptotiques proposées sont établies via une construction explicite des noyaux de Markov. Le Chapitre 6 est basé sur un article publié dans *Statistical Inference for Stochastic Processes.*

Mot clés: Expériences statistiques non paramétriques, distance de Le Cam, processus de diffusions, schéma d'Euler.

Abstract In Chapter 6 we establish the global asymptotic equivalence, in the sense of the Le Cam Δ-distance, between scalar diffusion models with unknown drift function and small variance on the one side, and nonparametric autoregressive models on the other side. The time horizon T is kept fixed and both the cases of discrete and continuous observation of the path are treated. We allow non constant diffusion coefficient, bounded but possibly tending to zero. The asymptotic equivalences are established by constructing explicit equivalence mappings. Chapter 6 is based on a paper published in *Statistical Inference for Stochastic Processes.*

Keywords: Nonparametric experiments, Le Cam distance, diffusion processes, Euler scheme.

6.1 Introduction

Diffusion processes obtained as small random perturbations of deterministic dynamical systems have been widely studied and have proved fruitful in applied problems (see e.g. Freidlin, Wentzell (2012)). Among other subjects, they have been applied to contingent claim pricing, see Uchida, Yoshida (2004) and the references therein, to filtering problems, see e.g. Picard (1986, 1991) and more recently to epidemic data, see Guy (2013). From a statistical point of view, these models have first been considered by Kutoyants (1984b) in the framework of continuous observation on a fixed time interval $[0,T]$. However, statistical inference for discretely observed diffusion processes has first been treated several years after, see Genon-Catalot (1990). In a nonparametric framework we may quote Kutoyants (1984a), among many others.

In this paper we consider the problem of estimating the drift function f associated with a scalar diffusion process (y_t) continuously or discretely observed on a time interval $[0,T]$, with $T<\infty$ kept fixed. More precisely, we consider the one-dimensional diffusion process (y_t) given by

$$dy_t = f(y_t)dt + \varepsilon\sigma(y_t)dW_t, \quad t\in[0,T], \quad y_0 = w\in\mathbb{R}, \tag{6.1}$$

where $(W_t)_{t\geq 0}$ is a standard $(\mathscr{A}_t)_{t\geq 0}$-Brownian motion defined on a probability space $(\Omega,\mathscr{A},\mathbb{P})$. The diffusion coefficient $\varepsilon\sigma(\cdot)$, with $0<\varepsilon<1$, is supposed to be known and to satisfy the following conditions:

(H1) $\sigma(\cdot)$ is a K-Lipschitz function on $\mathbb{R}$ bounded away from infinity and zero, i.e. there exist strictly positive constants σ_0,σ_1,K with

$$\sigma_0^2\leq\sigma^2(y)\leq\sigma_1^2 \ \text{ and } \ |\sigma(z)-\sigma(y)|\leq K|z-y|, \quad \forall z,y\in\mathbb{R}. \tag{6.2}$$

When (y_t) is discretely observed we will also require the following assumption:

(H2) $\sigma(\cdot)$ is a differentiable function on $\mathbb{R}$ with K-Lipschitz derivative, i.e.

$$|\sigma'(z)-\sigma'(y)|\leq K|z-y| \quad \forall z,y\in\mathbb{R}.$$

More in details, we consider two experiments, the continuous one associated with (y_t) and the discrete one given by the observations $(y_{t_1},\dots,y_{t_n})$, where $t_i=\frac{i}{n}T$. Our aim is to prove that these nonparametric experiments are both equivalent to an autoregressive model given by Euler type discretization of y with sampling interval $T/n, n \in \mathbb{N}^*$:

$$Z_0 = w, \quad Z_i = Z_{i-1} + \frac{T}{n} f(Z_{i-1}) + \varepsilon\sqrt{\frac{T}{n}}\sigma(Z_{i-1})\xi_i, \quad i = 1, \dots, n, \tag{6.3}$$

with independent standard normal variables ξ_i.

The concept of asymptotic equivalence that we shall adopt is based on the Le Cam Δ-distance between statistical experiments. Roughly speaking, saying that two statistical models, or experiments, are equivalent in the Le Cam sense means that any statistical inference procedure can be transferred from one model to the other in such a way that the asymptotic risk remains the same, at least for bounded loss functions. One can use this property in order to obtain asymptotic results working in a simpler but equivalent setting. For the basic concepts and a detailed description of the notion of asymptotic equivalence, we refer to Le Cam (1986); Le Cam, Yang (2000). A short review of this topic will be given in the Appendix.

In recent years, the Le Cam theory on the asymptotic equivalence between statistical models has aroused great interest and a large number of works has been published on this subject. In parametric statistics, Le Cam's theory has successfully been applied to a huge variety of experiments. Proving an asymptotic equivalence for nonparametric experiments is more demanding but, nowadays, several works in this subject have appeared. The first results of global asymptotic equivalence for nonparametric experiments date from 1996 and are due to Brown, Low (1996) and Nussbaum (1996). A non-exhausting list of subsequent works in this domain includes Brown et al. (2002b); Carter (2006b, 2007, 2009); Grama, Nussbaum (2002); Meister, Reiß (2013); Reiß (2008); Rohde (2004) for nonparametric regression, Brown et al. (2004a); Carter (2002); Jähnisch, Nussbaum (2003) for nonparametric density estimation models, Grama, Nussbaum (1998) for generalized linear models, Grama, Neumann (2006) for time series, Buchmann, Müller (2012) for GARCH model, Meister (2011) for functional linear regression, Golubev, Nussbaum, Zhou (2010) for spectral density estimation and Mariucci (2015b) for inhomogeneous jumps diffusion models. Negative results are somewhat harder to come by; the most notable ones among them are Brown, Zhang (1998); Efromovich, Samarov (1996); Wang (2002a).

Asymptotic equivalence results have also been obtained for diffusion models. References concern nonparametric drift estimation with known diffusion coefficient. Among these one can quote Dalalyan, Reiß (2006, 2007b); Delattre, Hoffmann (2002); Genon-Catalot, Laredo, Nussbaum (2002); Reiß (2011). However, the most relevant results to our purposes are due to Milstein, Nussbaum (1998) and to Genon-Catalot, Laredo (2014). The former authors have shown the asymptotic equivalence of a diffusion process continuously observed until time $T = 1$ having unknown drift function and constant small known diffusion coefficient, with the corresponding Euler scheme. They also proved the asymptotic sufficiency of the discretized observation of the diffusion with small sampling interval. Hence, our work is a generalization of Milstein, Nussbaum (1998). It can also be seen as a complement to Genon-Catalot, Laredo (2014), the difference being that in our case the time horizon is kept fixed and the diffusion coefficient goes to zero. This setting allows for weaker hypotheses than those assumed by Genon-Catalot and Larédo (for example, we do not need the drift function f to be uniformly bounded).

The interest in proving the asymptotic equivalence between the statistical model associated with the discretization of (6.1) and (6.3) lies in the difficulty of making inferences in the discretely observed diffusion model. On the other hand, inference for model (6.3) is well understood and in practice one often reduces to working with the latter (see e.g. Comte, Genon-Catalot, Rozenholc (2007); Genon-Catalot (1990); Hoffmann (1999); Laredo (1990)). The result in the present paper can thus be seen as a theoretical justification for such a practice.

The scheme of the proof is to prove both an asymptotic equivalence between the continuous and the discrete observation of (6.1) and one between the continuous model (6.1) and the Euler scheme (6.3). By the triangular inequality, the result will follow. The main difficulty lies in the model (6.3) being equivalent to a diffusion process with a diffusion coefficient $\bar{\sigma}$ different from σ. In particular, this means that the total variation distance between (6.3) and (6.1) is always 1. Thus, to prove the equivalence between these models it is necessary to construct an appropriate randomization. This is made possible by using random time changed experiments. Indeed, one can use random time changes in order to reduce to new diffusion models with diffusion coefficient equal to ε. However, these randomizations do not allow to apply the result of Milstein and Nussbaum directly since the changes of clock oblige to observe the new diffusion processes until different random times. Some care is then needed to overcome this technical obstacle (see Lemma 6.4.10).

The paper is organized as follows: In Section 2 we give a brief presentation of the most relevant references connected with our work. Section 3 contains the statement of the main results and a discussion while Section 4 is devoted to the proofs. The Appendix is devoted to background material.

6.2 Existing literature

As it has already been highlighted in the introduction, our result is not the first contribution in the context of asymptotic equivalences for diffusion processes. The aim of this section is to present the most relevant references linked with our work, that is Milstein, Nussbaum (1998), Genon-Catalot, Laredo (2014) and Dalalyan, Reiß (2006). We recall below the results contained in these papers.

- *Diffusion approximation for nonparametric autoregression*, Milstein, Nussbaum (1998): The authors consider the problem of estimating the function f from a continuously or discretely observed diffusion process $y(t)$, $t \in [0,1]$, which satisfies the SDE

  $$dy_t = f(y_t)dt + \varepsilon dW_t, \quad t \in [0,1], \quad y_0 = 0,$$

  where (W_t) is a standard Brownian motion and ε is a known small parameter.

  The drift function $f(\cdot)$ is unknown and such that, for K a positive constant,

  $$f \in \mathscr{F}_K = \Big\{ f \text{ defined on } \mathbb{R} \text{ and } \forall x, u \in \mathbb{R},\ |f(x) - f(u)| \leq K|x-u|,\ |f(0)| \leq K \Big\}.$$

  The constant K has to exist but may be unknown. For what concerns the discrete observation of (y_t) the authors place themselves in a high-frequency framework: $t_i = \frac{i}{n}$, $i \leq n$. Their main result is that, if $n\varepsilon \to \infty$ as $\varepsilon \to 0$, then there is an asymptotic equivalence between the continuous observation of (y_t) and the corresponding Euler scheme:

  $$Z_0 = 0, \quad Z_i = Z_{i-1} + \frac{f(Z_{i-1})}{n} + \frac{\varepsilon}{\sqrt{n}}\xi_i, \quad i = 1, \dots, n,$$

  where (ξ_i) are i.i.d. standard normal variables. Denoting by $\mathscr{P}$ and $\mathscr{Z}_n$ the statistical models associated with the continuous observation of (y_t) and the Euler scheme, respectively, an upper bound for the rate of convergence of $\Delta(\mathscr{P}, \mathscr{Z}_n)$ is given by:

  $$\Delta(\mathscr{P}, \mathscr{Z}_n) \leq O\Big(\sqrt{n^{-2}\varepsilon^{-2} + n^{-1}}\Big), \quad \text{as } \varepsilon \to 0.$$

  The authors also prove that the discrete observations $(y_{t_1}, \dots, y_{t_n})$ form an asymptotically sufficient statistics.

- *Asymptotic equivalence of nonparametric diffusion and Euler scheme experiments*, Genon-Catalot, Laredo (2014):

The authors consider the diffusion process (ξ_t) given by

$$d\xi_t = b(\xi_t)dt + \sigma(\xi_t)dW_t, \quad \xi_0 = \eta, \tag{6.4}$$

where (W_t) is a standard Brownian motion defined on a filtered probability space $(\Omega, \mathscr{A}, (\mathscr{A}_t)_{t\geq 0}, \mathbb{P})$, η is a real valued random variable, $\mathscr{A}_0$-measurable, $b(\cdot)$, $\sigma(\cdot)$ are real-valued functions defined on $\mathbb{R}$. The diffusion coefficient $\sigma(\cdot)$ is a known nonconstant function that belongs to $C^2(\mathbb{R})$ and satisfies the conditions:

$$\forall x \in \mathbb{R},\ 0 < \sigma_0^2 \leq \sigma^2(x) \leq \sigma_1^2, \quad |\sigma'(x)| + |\sigma''(x)| \leq K_\sigma.$$

The drift function $b(\cdot)$ is unknown and such that, for K a positive constant,

$$b(\cdot) \in \mathscr{F}_K = \Big\{b(\cdot) \in C^1(\mathbb{R}) \text{ and for all } x \in \mathbb{R},\ |b(x)| + |b'(x)| \leq K\Big\}.$$

The constant K has to exist but may be unknown. The sample path of (ξ_t) is continuously and discretely observed on a time interval $[0,T]$. The discrete observations of (ξ_t) occur at the times $t_i = ih$, $i \leq n$ with $T = nh$. The authors prove the asymptotic equivalence between the continuous or discrete observation of (ξ_t) and the corresponding Euler scheme:

$$Z_0 = \eta, \quad Z_i = Z_{i-1} + hb(Z_{i-1}) + \sqrt{h}\sigma(Z_{i-1})\varepsilon_i,$$

where, for $i \geq 1$, $t_i = ih$ and $\varepsilon_i = \frac{W_{t_i} - W_{t_{i-1}}}{\sqrt{h}}$. The equivalences hold under the assumptions that n tends to infinity with $h = h_n$ and $nh_n^2 = \frac{T^2}{n}$ tending to zero. This includes both the case $T = nh_n$ bounded and the one $T \to \infty$. Let us stress the rate of convergence in the small variance case, that is obtained by replacing σ by $\varepsilon\sigma$ in (6.4). Let us also denote by $\mathscr{P}$ and $\mathscr{Z}_n$ the statistical models associated with the continuous observation of (ξ_t) and the Euler scheme, respectively. The computations in Genon-Catalot, Laredo (2014) give the following upper bound for the rate of convergence:

$$\Delta(\mathscr{P}, \mathscr{Z}_n) \leq O\Big(\sqrt{n^{-2}\varepsilon^{-2} + n^{-1} + n^{-1}\varepsilon^{-4}}\Big).$$

- *Asymptotic statistical equivalence for scalar ergodic diffusions*, Dalalyan, Reiß (2006): The authors focus on diffusions processes of the form

$$dX_t = b(X_t)dt + dW_t, \quad t \in [0,T].$$

For some fixed constants $C, A, \gamma > 0$ the authors consider the non-parametric drift class

$$b \in \Sigma := \Big\{ b \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}) : |b(x)| \leq C(1+|x|), \forall |x| > A : b(x)\frac{x}{|x|} \leq -\gamma \Big\},$$

where $\mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})$ denotes the set of locally Lipschitz continuous functions. The class Σ has been chosen in order to ensure the existence of a stationary solution, unique in law, with invariant marginal density

$$\mu_b(x) = C_b \exp\Big(2\int_0^x b(y)dy\Big), \quad x \in \mathbb{R},$$

where $C_b > 0$ is a normalizing constant. The main result of the paper is the asymptotic equivalence between the model associated with the continuous observation of (X_t) and a certain Gaussian shift model, which can be interpreted as a regression model with random design.

However, in Section 4.5 the authors also consider discrete (high frequency) observations $(X_{t_i})_{i=0}^n$, where $0 = t_0 < t_1 < \cdots < t_n = T$, $d_i = t_i - t_{i-1}$ and $d_T = \max_{i=0,\dots,n-1} d_i$ goes to zero as T goes to infinity. It is shown the asymptotic sufficiency of $(X_{t_0}, \dots, X_{t_n})$ for $(X_t)_{t\in[0,T]}$ and the equivalence between the continuous observation of X and its discrete counterpart, that is the autoregression model defined by observing $(y_1, \dots, y_n)$ from

$$y_{i+1} = y_i + d_i b(y_i) + \sqrt{d_i}\xi_i, \quad i = 0, \dots, n-1, \quad y_0 \sim \mu_b,$$

where the ξ_i's are i.i.d. standard normal variables and independent of y_0.

6.3 Main results

To formulate our results we need to assume the standard conditions for existence and uniqueness of a strong solution y for the SDE (6.1) (Øksendal (1985), Theorem 5.5, page 45). We shall thus work with parameter spaces included in $\mathscr{F}_M$, the set of all functions f defined on $\mathbb{R}$ and satisfying

$$|f(0)| \leq M \text{ and } |f(z) - f(y)| \leq M|z-y|, \quad \forall z, y \in \mathbb{R}. \tag{6.5}$$

In particular, observe that every element of $\mathscr{F}_M$ satisfies a condition of linear growth: $|f(z)| \leq M(1+|z|)$, $\forall z \in \mathbb{R}$. Let $C = C(\mathbb{R}^+, \mathbb{R})$ be the space of continuous mappings ω from $\mathbb{R}^+$ into $\mathbb{R}$. Define the *canonical process* $x : C \to C$ by

$$\forall \omega \in C, \quad x_t(\omega) = \omega_t, \ \ \forall t \geq 0.$$

Let $\mathscr{C}^0$ be the smallest σ-algebra of parts of C that makes x_s, $s \geq 0$, measurable. Further, for any $t \geq 0$, let $\mathscr{C}_t^0$ be the smallest σ-algebra that makes x_s, s in $[0,t]$, measurable. Finally, set $\mathscr{C}_t := \bigcap_{s>t} \mathscr{C}_s^0$ and $\mathscr{C} := \sigma\big(\mathscr{C}_t; t \geq 0\big)$. Let us denote by $P_f^{n,y}$ the distribution induced on $(C, \mathscr{C}_T)$ by the law of y, solution to (6.1) and by $Q_f^{n,y}$ the distribution defined on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ by the law of $(y_{t_1}, \dots, y_{t_n})$, $t_i = T\frac{i}{n}$. We call $\mathscr{P}_y^T$ the experiment associated with the continuous observation of y until the time T and $\mathscr{Q}_y^n$ the discrete one, based on the grid values of y:

$$\mathscr{P}_y^T = \big(C, \mathscr{C}_T, \{P_f^y, f \in \mathscr{F}\}\big), \tag{6.6}$$

$$\mathscr{Q}_y^n = \big(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \{Q_f^{n,y}, f \in \mathscr{F}\}\big). \tag{6.7}$$

Finally, let us consider the experiment associated with the Euler scheme corresponding to (6.1). We denote by $Q_f^{n,Z}$ the distribution of $(Z_i, i = 1, \dots, n)$ defined by (6.3). Then:

$$\mathscr{Q}_Z^n = \big(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \{Q_f^{n,Z}, f \in \mathscr{F}\}\big). \tag{6.8}$$

Let us now state our main results.

Theorem 6.3.1. *Suppose that for some $M > 0$ the parameter space $\mathscr{F}$ fulfills $\mathscr{F} \subset \mathscr{F}_M$ and that $\sigma(\cdot)$ satisfies Assumption (H1) with $K = M$. Then, if $\varepsilon n \to \infty$ as $n \to \infty$ and $\varepsilon \to 0$, the experiments $\mathscr{P}_y^T$ and $\mathscr{Q}_Z^n$ are asymptotically equivalent. More precisely we have*

$$\Delta\big(\mathscr{P}_y^T, \mathscr{Q}_Z^n\big) = O\Big(\frac{1}{\varepsilon n} + (n^{-1} + \varepsilon)^{1/4}\Big).$$

Theorem 6.3.2. *Suppose that for some $M > 0$ the parameter space $\mathscr{F}$ fulfills $\mathscr{F} \subset \mathscr{F}_M$ and that $\sigma(\cdot)$ satisfies assumptions (H1) and (H2), with $K = M$. Furthermore, require that $\sigma(\cdot)$ and $\mathscr{F}$ are such that $\frac{f}{\sigma}(\cdot)$ is L-Lipschitz with a uniform L for $f \in \mathscr{F}$. Then, for any (possibly fixed) ε, the sampled values $y_{t_1}, \dots, y_{t_n}$ are an asymptotically sufficient statistic for the experiment $\mathscr{P}_y^T$.*

Remark 6.3.3. If $\mathscr{F} \subset \mathscr{F}_M$ for some $M > 0$ and $\sigma(\cdot)$ satisfies assumptions (H1) and (H2) then the Lipschitz condition on $\frac{f}{\sigma}(\cdot)$ is satisfied in both the following two cases: Either $|f(x)| \leq M$ for all x, or $|x\sigma'(x)| \leq M$ for all x.

Corollary 6.3.4. *Under the same hypotheses as in Theorem 6.3.2, the statistical model associated with the sampled values $y_{t_1}, \dots, y_{t_n}$ is asymptotically equivalent to $\mathscr{Q}_Z^n$, as n goes to infinity. The same upper bound for the rate of convergence that appears in Theorem 6.3.1 holds.*

6.3.1 Discussion

Our results are intended to be a generalization of Milstein, Nussbaum (1998) allowing to have a small but non-constant diffusion coefficient.

The fact of being in a small variance case is crucial for the proof of our results. Indeed, as in Genon-Catalot, Laredo (2014), we use auxiliary random time changed models to prove Theorem 6.3.1 and a key step in the proof is the comparison between diffusion models having small (and constant) diffusion coefficient ε observed until different stopping times. In particular, this leads to compare the L_1-norm between two stopping times (see Lemma 6.4.10 as opposed to Lemma 3.2 in Genon-Catalot, Laredo (2014)) and we take care of that by using the fact that any diffusion process with small variance converges to some deterministic solution. As a consequence, we find the same conditions as in Milstein, Nussbaum (1998) on ε and n, i.e. $n\varepsilon \to \infty$, instead of conditions of the type $n\varepsilon^4 \to \infty$ as in Genon-Catalot, Laredo (2014). Another important feature of the small variance case is that it allows for weaker hypotheses than those assumed in Genon-Catalot, Laredo (2014); more precisely, we only need to ask the same conditions about $\mathscr{F}$ as in Milstein, Nussbaum (1998).

Also remark that the most important novelty in the paper is Theorem 6.3.1 since Theorem 6.3.2 is a small generalization of the results already obtained in Genon-Catalot, Laredo (2014), Dalalyan, Reiß (2006).

6.4 Proofs

In this section we collect the proofs of Theorems 6.3.1 and 6.3.2. Since a lot of auxiliary statistical models are used to obtain our main results, we believe that starting by outlining the strategy of the proofs can be helpful to the reader. More precisely, seven models come into play in the proof of Theorem 6.3.1:

$(\mathscr{M}_1)$ $dy_t = f(y_t)dt + \varepsilon\sigma(y_t)dW_t, \quad y_0 = w, \quad t \in [0,T];$

$(\mathscr{M}_2)$ $d\bar{y}_t = \bar{f}_n(t,\bar{y})dt + \varepsilon\bar{\sigma}_n(t,\bar{y})dW_t, \quad \bar{y}_0 = w, \quad t \in [0,T];$

$(\mathscr{M}_3)$ $d\xi_t = \frac{f}{\sigma^2}(\xi_t)dt + \varepsilon dW_t, \quad \xi_0 = w, \quad t \in [0, A_T(\xi)];$

$(\mathscr{M}_4)$ $d\bar{\xi}_t = \bar{g}_n(t,\bar{\xi})dt + \varepsilon dW_t, \quad \bar{\xi}_0 = w, \quad t \in [0, \bar{A}_T^n(\bar{\xi})];$

$(\mathscr{M}_5)$ $d\xi_t = \frac{f}{\sigma^2}(\xi_t)dt + \varepsilon dW_t, \quad \xi_0 = w, \quad t \in [0, S_T^n(\xi)];$

$(\mathscr{M}_6)$ $d\xi_t = \frac{f}{\sigma^2}(\xi_t)dt + \varepsilon dW_t, \quad \xi_0 = w, \quad t \in [0, \bar{A}_T(\bar{\xi})];$

($\mathscr{M}_7$) $Z_0 = w, \quad Z_i = Z_{i-1} + \frac{T}{n} f(Z_{i-1}) + \varepsilon\sqrt{\frac{T}{n}}\sigma(Z_{i-1})\xi_i, \quad i = 1, \dots, n;$

where $A_T(x)$, $\bar{A}_T^n(x)$ and $S_T^n(x)$ are certain $\mathscr{C}_t$-stopping times; $\bar{f}_n$, $\bar{\sigma}_n$, $\bar{g}_n$ are piecewise constant approximations of f and σ, and the ξ_i's are independent standard normal variables.

The scheme of the proof is as follows:

- $\Delta(\mathscr{M}_1, \mathscr{M}_3)$, $\Delta(\mathscr{M}_2, \mathscr{M}_4)$ and $\Delta(\mathscr{M}_2, \mathscr{M}_7)$ are equal to zero: see, respectively, Propositions 6.4.3, 6.4.5 and 6.4.12.
- $\Delta(\mathscr{M}_5, \mathscr{M}_3)$ and $\Delta(\mathscr{M}_5, \mathscr{M}_6)$ are bounded by $\sqrt[4]{n^{-1}+\varepsilon}$ up to some constants: see Proposition 6.4.11
- $\Delta(\mathscr{M}_6, \mathscr{M}_4) = O\Big(\sqrt{n^{-1}+\varepsilon^{-2}n^{-2}}\Big)$: see Proposition 6.4.9.

We thus deduce that the Le Cam Δ-distance between our models of interest is bounded by:

$$\Delta(\mathscr{M}_1, \mathscr{M}_7) \leq O\Big(\sqrt[4]{n^{-1}+\varepsilon} + \varepsilon^{-1}n^{-1}\Big).$$

On the other hand, six the models are used in the Proof of Theorem 6.3.2:

($\mathscr{N}_1$) $dy_t = f(y_t)dt + \varepsilon\sigma(y_t)dW_t, \quad y_0 = w, \quad t \in [0, T];$

($\mathscr{N}_2$) $(y_{t_1}, \dots, y_{t_n});$

($\mathscr{N}_3$) $d\mu_t = \Big(\frac{f(F^{-1}(\mu_t))}{\varepsilon\sigma(F^{-1}(\mu_t))} - \frac{\sigma'(F^{-1}(\mu_t))}{2\varepsilon}\Big)dt + dW_t, \quad \mu_0 = F(w), \quad t \in [0, T];$

($\mathscr{N}_4$) $(\mu_{t_1}, \dots, \mu_{t_n});$

($\mathscr{N}_5$) $d\bar{\mu}_t = \bar{b}_n(t, \bar{\mu})dt + dW_t, \quad \bar{\mu}_0 = F(w), \quad t \in [0, T];$

($\mathscr{N}_6$) $(\bar{\mu}_{t_1}, \dots, \bar{\mu}_{t_n});$

where $F(x) = \int_0^x \frac{1}{\varepsilon\sigma(u)}du$ and $\bar{b}_n$ is a piecewise constant approximation of a certain function b depending on f, ε, σ and F.

The strategy of the proof is:

- $\Delta(\mathscr{N}_1, \mathscr{N}_3)$, $\Delta(\mathscr{N}_2, \mathscr{N}_4)$ and $\Delta(\mathscr{N}_5, \mathscr{N}_6)$ are equal to zero: see Propositions 6.4.13 and 6.4.15;
- $\Delta(\mathscr{N}_3, \mathscr{N}_5)$ and $\Delta(\mathscr{N}_6, \mathscr{N}_4)$ are bounded by n^{-1} up to some constants: see Propositions 6.4.14 and 6.4.15, respectively.

It follows that:

$$\Delta(\mathscr{N}_1, \mathscr{N}_2) = O\big(n^{-1}\big).$$

6.4.1 Random time substitutions for Markov processes

A key tool in establishing the asymptotic equivalence between the diffusion model continuously observed and its Euler scheme is given by random time changes for Markov processes. More in details we will need the following results.

Theorem 6.4.1. *(see Volkonskii (1958)) Let $(Y, \mathbb{P}_y)$ be a (càdlàg) strong $(\mathscr{A}_t)$-Markov process on $(\Omega, \mathscr{A}, \mathbb{P})$ with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and let $v : \mathbb{R}^d \to (0, \infty)$ be a positive continuous function. Define the additive functional*

$$F_t = \int_0^t \frac{ds}{v(Y_s)},\ t \geq 0$$

and assume that

$$\int_0^\infty \frac{ds}{v(Y_s)} = \infty,\ \mathbb{P}_y - \text{a.s. } ,\forall y \in \mathbb{R}^d,$$

so that the right continuous inverse

$$T_t = \inf\{s \geq 0 : F_s > t\},\ t \geq 0$$

of the functional F is well defined on $[0, \infty)$. Then the process

$$J_t = Y_{T_t},\ t \geq 0,$$

is a càdlàg strong $(\mathscr{A}_{T_t})$-Markov process on the probability space $(\Omega, \mathscr{A}, \mathbb{P})$.

Assume moreover that $(Y, \mathbb{P}_y)$ is a Feller process with infinitesimal generator $\mathcal{L}^Y$ and domain $\mathcal{D}$. Then J is also a Feller process whose infinitesimal generator, with domain $\mathcal{D}$, is given by

$$\mathcal{L}^J h(z) = v(z)\mathcal{L}^Y h(z), \quad h \in \mathcal{D},\ z \in \mathbb{R}^d.$$

Property 6.4.2. *For all $\omega \in C$, $s, t > 0$ define:*

$$\rho_s(\omega) = \int_0^s \sigma^2(\omega_r)dr; \quad \eta_t(\omega) = \inf\big\{s \geq 0,\ \rho_s(\omega) \geq t\big\},$$
$$\theta_s(\omega) = \int_0^s \frac{1}{\sigma^2(\omega_r)}dr; \quad A_t(\omega) = \inf\big\{s \geq 0,\ \theta_s(\omega) \geq t\big\}.$$

Then, the following hold:

1. $\rho_T(x) = A_T(x_{\eta_\cdot(x)})$,
2. $A_t(x) = \int_0^t \sigma^2(x_{A_s(x)})ds, \quad \forall t \in [0, T]$.

Proof. 1. It is enough to show that $\theta_T(x_{\eta_\cdot(x)}) = \eta_T^n(x)$ since $t \mapsto A_t(x)$ and $t \mapsto \rho_t(x)$ are, respectively, the inverses of the applications $t \mapsto \theta_t(x)$ and $t \mapsto \eta_t(x)$. To prove the last assertion compute:

$$\theta_T(x_{\eta_\cdot(x)}) = \int_0^T \frac{1}{\sigma^2(x_{\eta_r(x)})} dr = \int_0^{\eta_T(x)} \frac{\sigma^2(x_s)}{\sigma^2(x_s)} ds = \eta_T(x);$$

where in the second equality we have performed the change of variable $s = \eta_r(x) \Leftrightarrow r = \rho_s(x)$ that yields $dr = \sigma^2(x_s)ds$.

2. Again, we use that $t \mapsto \theta_t(x)$ is the inverse of the function $t \mapsto A_t(x)$ combined with the following elementary fact:

Let h and g be two differentiable functions on $\mathbb{R}$ such that $h(0) = 0 = g(0)$ and their derivatives never vanish. Then, $h'(z) = \frac{1}{g'(h(z))}$ for all z in $\mathbb{R}$ if and only if h is the inverse of g.

To show the assertion in 2. it is enough to apply this fact to $h(t) = A_t(x)$ and $g(t) = \theta_t(x)$. □

6.4.2 Proof of Theorem 6.3.1

We will proceed in four steps. More precisely, in Step 1 we consider a random time change on the diffusion (6.1) in order to obtain an experiment equivalent to $\mathscr{P}_y^T$ but associated with a diffusion having diffusion coefficient equal to ε. In Step 2 we construct a continuous time discretization of the process (y_t) and, applying a second random time change, we prove an equivalence result between a second experiment associated again with a diffusion having diffusion coefficient equal to ε. In Step 3 we compare, in term of the Le Cam Δ-distance, the two experiments with the diffusion coefficient equal to ε constructed in Steps 1-2. Finally, in Step 4, we prove the equivalence between the experiment associated with the continuous time discretization of (y_t) and the one with the Euler scheme. By means of the triangular inequality we are able to bound the Le Cam Δ-distance between $\mathscr{P}_y^T$ and $\mathscr{Q}_Z^n$.

Step 1. We start by proving the Le Cam equivalence between $\mathscr{P}_y^T$ and a corresponding diffusion model with coefficient diffusion equal to ε. Recall that P_f^y is the law on $(C, \mathscr{C}_T)$ of a diffusion process with infinitesimal generator $\mathcal{L}_1$ given by

$$\mathcal{L}_1 = f\nabla + \varepsilon^2 \frac{\sigma^2}{2}\Delta, \tag{6.9}$$

Define P_f^ξ as the law on $(C, \mathscr{C})$ of a diffusion process with infinitesimal generator $\mathcal{L}_2$ given by

$$\mathcal{L}_2 = \frac{f}{\sigma^2}\nabla + \varepsilon^2 \frac{1}{2}\Delta \tag{6.10}$$

and initial condition $\xi_0 = \omega$. Moreover, for all $A > 0$ a $\mathscr{C}_t$-stopping time, define the experiment

$$\mathscr{P}_\xi^A = (C, \mathscr{C}_A, (P_f^\xi|_{\mathscr{C}_A}, f \in \mathscr{F})).$$

Proposition 6.4.3. *Suppose that for some $M > 0$ the parameter space $\mathscr{F}$ fulfills $\mathscr{F} \subset \mathscr{F}_M$ and that $\sigma(\cdot)$ satisfies assumption (H1), with $K = M$. Then, the statistical models $\mathscr{P}_y^T$ and $\mathscr{P}_\xi^{A_T(x)}$ are equivalent.*

Proof. Let us prove that $\delta(\mathscr{P}_y^T, \mathscr{P}_\xi^{A_T(x)}) = 0$. Note that (x_t) under P_f^y is a $(\mathscr{C}_t)$-Markov process with infinitesimal generator as in (6.9). Define a new process ξ as a change of time of (x_t) with stochastic clock $(\eta_t(x))_t$: $\xi_0 = w$, $\xi_t := x_{\eta_t(x)}$, $\forall t > 0$. Theorem 6.4.1 ensures that the process $(\xi_t)_{t\geq 0}$ is a diffusion process with infinitesimal generator given by (6.10). Also, remark that, as (x_t) is defined on $[0,T]$, then the trajectories of (ξ_t) are defined until the time $\rho_T(x)$. In order to produce a randomization transforming the family of measures $\{P_f^y,\ f \in \mathscr{F}\}$ in $\{P_f^\xi|\mathscr{C}_{A_T(x)},\ f \in \mathscr{F}\}$, let us consider the following application:

$$\Phi : \{\omega_t :\ t \in [0,T]\} \to \{\omega_{\eta_t(\omega)} :\ t \in [0, \rho_T(\omega)]\}.$$

Observe that the process $\Phi(x)$ is defined until the time $\rho_T(x)$ that is equal to $A_T(\Phi(x))$ (see Property 6.4.2), so that any set of paths of $\Phi(x)$ belongs to $\mathscr{C}_{A_T(x)}$. Introduce the Markov kernel K defined by $K(\omega, \Gamma) = \mathbb{I}_\Gamma(\Phi(\omega))$, $\forall \omega \in C$, $\forall \Gamma \in \mathscr{C}_{A_T(x)}$, then:

$$KP_f^y(\Gamma) = \int \mathbb{I}_\Gamma(\Phi(\omega)) P_f^y(d\omega) = P_f^y(\Phi(x) \in \Gamma) = P_f^\xi|_{\mathscr{C}_{A_T(x)}}(\Gamma).$$

Therefore $\delta(\mathscr{P}_y^T, \mathscr{P}_\xi^{A_T(x)}) = 0$.

The same type of computations imply that $\delta(\mathscr{P}_\xi^{A_T(x)}, \mathscr{P}_y^T) = 0$ through use of the application $\Psi : (\omega_t :\ t \in [0, A_T(\omega)]) \to (\omega_{A_t(\omega)} :\ t \in [0,T])$. □

Step 2. We now introduce a statistical model that approximates the model $\mathscr{P}_y^T$. Given a path ω in C and a time grid $t_i = T\frac{i}{n}$, we define

$$\bar{f}_n(t,\omega) = \sum_{i=1}^{n-1} f\big(\omega(t_i)\big)\mathbb{I}_{[t_i,t_{i+1})}(t), \quad \bar{\sigma}_n(t,\omega) = \sum_{i=1}^{n-1} \sigma\big(\omega(t_i)\big)\mathbb{I}_{[t_i,t_{i+1})}(t), \quad \forall t \in [0,T].$$

Then, we denote by $P_f^{n,\bar{y}}$ the law on $(C, \mathscr{C}_T)$ of a diffusion process with infinitesimal generator $\bar{\mathcal{L}}^n$ given by

$$\bar{\mathcal{L}}_t^n(\omega)h(z) = \bar{f}_n(t,\omega)\nabla h(z) + \varepsilon^2 \frac{\bar{\sigma}_n^2(t,\omega)}{2}\Delta h(z), \quad \forall \omega \in C,\ h \in C^2(\mathbb{R}),\ z \in \mathbb{R} \tag{6.11}$$

and initial condition $\bar{y}_0 = \omega$. Consider the experiment

$$\mathscr{P}_{\bar{y}}^{n,T} = \big(C, \mathscr{C}_T, (P_f^{n,\bar{y}}|_{\mathscr{C}_T}, f \in \mathscr{F})\big)$$

Again, we want to introduce the diffusion model with diffusion coefficient equal to ε associated to $\mathscr{P}_{\bar{y}}^{n,T}$. To that aim, for all $\omega \in C$, define

$$\bar{A}_0^n(\omega) = 0, \quad \bar{A}_t^n(\omega) = \bar{A}_{t_{i-1}}(\omega) + \sigma^2(\omega_{\bar{A}_{t_{i-1}}}(\omega))(t - t_{i-1}), \quad t \in (t_{i-1}, t_i]; \tag{6.12}$$

$$\bar{g}_n(t,\omega) = \sum_{i=1}^n \frac{f}{\sigma^2}(\omega_{\bar{A}_{t_i}^n(\omega)})\mathbb{I}_{\left(\bar{A}_{t_i}^n(\omega), \bar{A}_{t_{i+1}}^n(\omega)\right]}(t), \quad t \geq 0, \quad i = 0, \dots, n-1. \tag{6.13}$$

Lemma 6.4.4. *$\bar{A}_{t_i}^n(x)$ is a $\mathscr{C}_t$-stopping time for all $i = 1, \dots, n$.*

Proof. By (6.12), $\bar{A}_{t_1}^n(x) = \sigma^2(x(0))t_1$, so the set $\{\bar{A}_{t_1}^n(x) \leq t\} = \begin{cases} \emptyset & \text{if } \sigma^2(x(0))t_1 > t \\ C & \text{otherwise.} \end{cases}$ belongs to $\mathscr{C}_t$, for all t. By induction, assume that $\bar{A}_{t_{i-1}}^n(x)$ is a $(\mathscr{C}_t)$-stopping time and remark that (6.12) implies $\{\bar{A}_{t_i}^n(x) \leq t\} = \{\bar{A}_{t_i}^n(x) \leq t\} \cap \{\bar{A}_{t_{i-1}}^n(x) \leq t\}$. Since (x_t) is $(\mathscr{C}_t)$-adapted and continuous, in particular it is progressively measurable with respect to $(\mathscr{C}_t)$. By the induction hypothesis it follows that $x_{\bar{A}_{t_{i-1}}^n(x)}$ is $\mathscr{C}_{\bar{A}_{t_{i-1}}(x)}$- measurable, hence, using (6.12), $\{\bar{A}_{t_i}^n(x) \leq t\} \in \mathscr{C}_{\bar{A}_{t_{i-1}}^n(x)}$, as $\bar{A}_{t_{i-1}}^n(x)$ is already $\mathscr{C}_{\bar{A}_{t_{i-1}}^n(x)}$-measurable, again by the induction hypothesis. By the definition of the σ-algebra $\mathscr{C}_{\bar{A}_{t_{i-1}}(x)}$ and the induction hypothesis, we then conclude that $\{\bar{A}_{t_i}^n(x) \leq t\} \in \mathscr{C}_t$. Hence the result. □

Denote by $P_f^{n,\bar{\xi}}$ the law on $(C, \mathscr{C})$ of a diffusion process with infinitesimal generator $\tilde{\mathcal{L}}^n$ given by

$$\tilde{\mathcal{L}}_t^n(\omega)h(z) = \bar{g}_n(t,\omega)\nabla h(z) + \frac{\varepsilon^2}{2}\Delta h(z), \quad \forall \omega \in C, \; h \in C^2(\mathbb{R}), \; z \in \mathbb{R} \tag{6.14}$$

and initial condition $\bar{\xi}_0 = \omega$. Thanks to Lemma 6.4.4 we can define the statistical model associated with the observation of $\bar{\xi}$ until the stopping time $\bar{A}_T^n(x)$:

$$\mathscr{P}_{\bar{\xi}}^{n,\bar{A}_T^n(x)} = \big(C, \mathscr{C}_{\bar{A}_T^n(x)}, \{P_f^{n,\bar{\xi}}|_{\mathscr{C}_{\bar{A}_T^n(x)}}, f \in \mathscr{F}\}\big).$$

As in Step 1, one can prove the following proposition. There are, however, some technical points that need to be taken care of; for more details, we refer to Genon-Catalot, Laredo (2014), Proposition 5.4.

Proposition 6.4.5. *Under the same hypotheses as in Proposition 6.4.3, the statistical models $\mathscr{P}_{\bar{y}}^{n,T}$ and $\mathscr{P}_{\bar{\xi}}^{n,\bar{A}_T^n(x)}$ are equivalent.*

Step 3. We shall prove that $\Delta(\mathscr{P}_\xi^{A_T(x)}, \mathscr{P}_{\bar{\xi}}^{n,\bar{A}_T^n(x)}) \to 0$ as $n \to \infty$. To that aim we will prove that, setting $S_T^n(x) := A_T(x) \wedge \bar{A}_T^n(x)$, $\Delta(\mathscr{P}_\xi^{A_T(x)}, \mathscr{P}_\xi^{S_T^n(x)}) \to 0$, $\Delta(\mathscr{P}_\xi^{S_T^n(x)}, \mathscr{P}_\xi^{\bar{A}_T^n(x)}) \to 0$ and $\Delta(\mathscr{P}_\xi^{\bar{A}_T^n(x)}, \mathscr{P}_{\bar{\xi}}^{n,\bar{A}_T^n(x)}) \to 0$ as $n \to \infty$. We shall start by showing the asymptotic equivance between $\mathscr{P}_\xi^{\bar{A}_T^n(x)}$ and $\mathscr{P}_{\bar{\xi}}^{n,\bar{A}_T^n(x)}$; we need the following lemmas:

Lemma 6.4.6. *The law of $(x_{A_t(x)})$ under P_f^ξ is the same as the law of (x_t) under P_f^y. Moreover, let $P_f^{\bar{\zeta}}$ be the distribution induced on $(C, \mathscr{C})$ by the law of a diffusion process $(\bar{\zeta}_t)$ satisfying*

$$d\bar{\zeta}_t = \frac{f(\bar{\zeta}_t)}{\sigma^2(\bar{\zeta}_t)} \bar{\sigma}_n^2(t, \bar{\zeta}) dt + \varepsilon dW_t, \quad \bar{\zeta}_0 = w.$$

Then, the law of $(x_{\bar{A}_t^n(x)})$ under P_f^ξ is the same as the law of (x_t) under $P_f^{\bar{\zeta}}$.

Proof. We shall only prove the first assertion, the proof of the second one being very similar. We have $\xi_0 = w = y_0$ and, for all $t > 0$:

$$x_{A_t(x)} = w + \int_0^{A_t(x)} \frac{f(x_s)}{\sigma^2(x_s)} ds + \varepsilon \tilde{W}_{A_t^n(x)},$$

where the process $(\tilde{W}_t)$ is a standard Brownian motion under P_f^ξ. The change of variable $s = A_u(x)$ implies that $ds = \sigma^2(x_{A_u(x)}) du$, hence one can write

$$x_{A_t(x)} = w + \int_0^t \frac{f(x_{A_u(x)})}{\sigma^2(x_{A_u(x)})} \sigma^2(x_{A_u(x)}) du + \varepsilon \int_0^t \sigma(x_{A_u(x)}) B_u,$$

where the process (B_t) is defined by

$$B_t = \int_0^t \frac{d\tilde{W}_{A_u(x)}}{\sigma(x_{A_u(x)})}.$$

Classical results (see e.g. Karatzas, Shreve (2000), 5.5) ensure that (B_t) is a $\mathscr{C}_{A_t(x)}$ standard Brownian motion under P_f^y. It follows that the law of $x_{A_t(x)}$ under P_f^ξ is the same as the law of x under P_f^y. □

Lemma 6.4.7. *Let p be an even positive integer and (t_n) a sequence of times bounded by CT for some constant C independent of f; then $\mathbb{E}_{P_f^{\bar{\zeta}}} |x_{t_n}|^p = O(1)$, uniformly on $\mathscr{F}$.*

Proof. In order to bound $\mathbb{E}_{P_f^{\bar{\zeta}}} |x_{t_n}|^p$ we will use the following facts:

- $(z_1 + \cdots + z_m)^p \leq m^{p-1}(z_1^p + \cdots + z_m^p)$, $\forall z_1, \ldots, z_m \in \mathbb{R}$;

- $\mathbb{E}_{P_f^{\bar\zeta}}|\int_0^v h(x_r)dr|^p \leq v^{p-1}\int_0^v \mathbb{E}_{P_f^{\bar\zeta}} h^p(x_r)dr$, for any integrable function h;
- If X is a centered Gaussian random variable with variance σ^2 then $\mathbb{E}[X^p] = \sigma^p(p-1)!!$;
- (Gronwall lemma) Let $I = [0, a]$ be an interval of the real line, α a constant and let β and u continuous real valued functions defined on I. If β is non-negative and if u satisfies the integral inequality:
$$u(t) \leq \alpha + \int_0^t \beta(s)u(s)\,\mathrm{d}s, \qquad \forall t \in I,$$
then
$$u(t) \leq \alpha \exp\Big(\int_0^t \beta(s)\,\mathrm{d}s\Big), \qquad t \in I.$$

As one can always construct a Brownian motion $(\bar B_t)$ under $P_f^{\bar\zeta}$ such that
$$dx_t = \frac{f(x_t)}{\sigma^2(x_t)}\bar\sigma_n^2(t,x)dt + \varepsilon d\bar B_t;$$
applying the first three facts combined with the linear growth of f one can write:
$$\begin{aligned}
\mathbb{E}_{P_f^{\bar\zeta}}|x_{t_n}|^p &\leq 3^{p-1}w^p + 3^{p-1}\mathbb{E}_{P_f^{\bar\zeta}}\Big(\int_0^{t_n}\frac{f(x_s)}{\sigma^2(x_s)}\bar\sigma_n^2(s,x)ds\Big)^p + 3^{p-1}\varepsilon^p\mathbb{E}_{P_f^{\bar\zeta}}\bar B_{t_n}^p\\
&\leq 3^{p-1}w^p + 3^{p-1}\frac{\sigma_1^{2p}}{\sigma_0^{2p}}(CT)^{p-1}\int_0^{t_n}\mathbb{E}_{P_f^{\bar\zeta}}[f^p(x_s)]ds + 3^{p-1}\varepsilon^p(CT)^{\frac{p}{2}}(p-1)!!\\
&\leq C'\Big(1+\int_0^{t_n}\mathbb{E}_{P_f^{\bar\zeta}}|x_s|^p ds\Big),
\end{aligned}$$
for some constant C' independent of f. Applying the Gronwall lemma, we obtain
$$\mathbb{E}_{P_f^{\bar\zeta}}|x_{t_n}|^p \leq C'e^{C't_n} \leq C'e^{C'CT} = O(1).$$
$\square$

Lemma 6.4.8. *Under the same hypotheses as in Proposition 6.4.3 and with the same notation as in Steps 1 and 2, we have*
$$\mathbb{E}_{P_f^{\xi}}\int_0^{\bar A_T^n(x)}\Big(\frac{f(x_s)}{\sigma^2(x_s)} - \bar g_n(s,x)\Big)^2 ds = O\big(n^{-2}+\varepsilon n^{-1}\big),$$
uniformly on $\mathscr{F}$.

Proof. For the sake of brevity, in this proof we will omit the superscript n in each occurrence of $\bar{A}^n_t$. We start by observing that, for all $y, z \in \mathbb{R}$

$$\begin{aligned}\left|\frac{f(z)}{\sigma^2(z)} - \frac{f(y)}{\sigma^2(y)}\right| &\le |f(z)|\left|\frac{1}{\sigma^2(z)} - \frac{1}{\sigma^2(y)}\right| + \frac{|f(z)-f(y)|}{\sigma_0^2}\\ &\le \frac{2M^2\sigma_1}{\sigma_0^4}(1+|z|)|z-y| + \frac{M}{\sigma_0^2}|z-y|,\end{aligned}$$

hence there exists some constant C such that $\left(\frac{f(z)}{\sigma^2(z)} - \frac{f(y)}{\sigma^2(y)}\right)^2 \le C(z-y)^2(1+z^2)$.

Applying this inequality we can write:

$$\begin{aligned}\int_0^{\bar{A}_T(x)} \Big(\frac{f(x_s)}{\sigma^2(x_s)} - \bar{g}_n(s,x)\Big)^2 ds &\le \sum_{i=0}^{n-1} \int_{\bar{A}_{t_i}(x)}^{\bar{A}_{t_{i+1}}(x)} C\big(x_s - x_{\bar{A}_{t_i}(x)}\big)^2 (1 + x^2_{\bar{A}_{t_i}(x)}) ds\\ &= C\sum_{i=0}^{n-1}(1 + x^2_{\bar{A}_{t_i}(x)}) \int_{\bar{A}_{t_i}(x)}^{\bar{A}_{t_{i+1}}(x)} (x_r - x_{\bar{A}_{t_i}(x)})^2 dr\\ &\le C\sigma_1^2 \sum_{i=0}^{n-1}(1 + x^2_{\bar{A}_{t_i}(x)}) \int_0^{t_{i+1}-t_i} (x_{\bar{A}_{t_i+s}(x)} - x_{\bar{A}_{t_i}(x)})^2 ds,\end{aligned}$$

where in the last step we have performed the change of variables $r = \bar{A}_{t_i+s}(x)$. Thanks to the Cauchy-Schwarz inequality and Lemma 6.4.7 we obtain

$$\begin{aligned}&\mathbb{E}_{P_f^\xi} \int_0^{\bar{A}_T(x)} \Big(\frac{f(x_s)}{\sigma^2(x_s)} - \bar{g}_n(s,x)\Big)^2 ds \le\\ &\le C\sigma_1^2 \sum_{i=0}^{n-1} \sqrt{\mathbb{E}_{P_f^\xi}(1 + x^2_{\bar{A}_{t_i}(x)})^2} \sqrt{\mathbb{E}_{P_f^\xi}\Big(\int_0^{t_{i+1}-t_i} (x_{\bar{A}_{t_i+r}(x)} - x_{\bar{A}_{t_i}(x)})^2 dr\Big)^2}\\ &= C\sigma_1^2 \sum_{i=0}^{n-1} \sqrt{\mathbb{E}_{P_f^{\bar{\zeta}}}(1 + x^2_{t_i})^2} \sqrt{\mathbb{E}_{P_f^{\bar{\zeta}}}\Big(\int_0^{\frac{T}{n}} (x_{r+t_i} - x_{t_i})^2 dr\Big)^2}\\ &\le C\sigma_1^2 \sum_{i=0}^{n-1} \sqrt{\mathbb{E}_{P_f^{\bar{\zeta}}}(2 + 2x^4_{t_i})} \sqrt{\mathbb{E}_{P_f^{\bar{\zeta}}}\Big(\frac{T}{n}\int_0^{\frac{T}{n}} (x_{r+t_i} - x_{t_i})^4 dr\Big)}\\ &= O\Bigg(\sum_{i=0}^{n-1} \sqrt{\mathbb{E}_{P_f^{\bar{\zeta}}}\Big(\frac{T}{n}\int_0^{\frac{T}{n}} (x_{r+t_i} - x_{t_i})^4 dr\Big)}\Bigg).\end{aligned}$$

Using the same arguments as in the proof of Lemma 6.4.7, we can write

$$\begin{aligned}
\mathbb{E}_{P_f^{\bar\xi}}(x_{r+t_i}-x_{t_i})^4 &= \mathbb{E}_{P_f^{\bar\xi}}\bigg|\int_{t_i}^{t_i+r}\frac{f(x_s)\sigma^2(x_{t_i})}{\sigma^2(x_s)}ds+\varepsilon\sigma(x_{t_i})(\bar B_{t_i+r}-\bar B_{t_i})\bigg|^4\\
&\le 8\mathbb{E}_{P_f^{\bar\xi}}\bigg|\int_{t_i}^{t_i+r}\frac{f(x_s)\sigma^2(x_{t_i})}{\sigma^2(x_s)}ds\bigg|^4+8\varepsilon^4\sigma_1^4\mathbb{E}_{P_f^{\bar\xi}}\big|\bar B_{t_i+r}-\bar B_{t_i}\big|^4\\
&\le 8\frac{\sigma_1^8}{\sigma_0^8}r^3\int_0^r\mathbb{E}_{P_f^{\bar\xi}}\big[f^4(x_{s+t_i})\big]ds+8\varepsilon^4\sigma_1^4r^2 6!\\
&= O\bigg(r^4+r^3\int_0^r\mathbb{E}_{P_f^{\bar\xi}}[x_{s+t_i}^4]ds+\varepsilon^4\ r^2\bigg)\\
&= O(r^4+\varepsilon^4r^2)=O\Big(\frac{1}{n^4}+\frac{\varepsilon^4}{n^2}\Big).
\end{aligned}$$

Putting all the pieces together we get:

$$\begin{aligned}
\int_0^{\bar A_T(x)}\Big(\frac{f(x_s)}{\sigma^2(x_s)}-\bar g_n(s,x)\Big)^2ds &= O\bigg(\sum_{i=0}^{n-1}\sqrt{\Big(\frac{T}{n}\int_0^{\frac{T}{n}}O\Big(\frac{1}{n^4}+\frac{\varepsilon^4}{n^2}\Big)dr\Big)}\bigg)\\
&= O\bigg(\frac{1}{n^2}+\frac{\varepsilon^2}{n}\bigg).
\end{aligned}$$

□

Proposition 6.4.9. *Under the same hypotheses as in Proposition 6.4.3, we have*

$$\Delta(\mathscr{P}_\xi^{\bar A_T^n(x)},\mathscr{P}_{\bar\xi}^{n,\bar A_T^n(x)})=O(\sqrt{\varepsilon^{-2}n^{-2}+n^{-1}}).$$

Proof. We use an inequality involving the Hellinger process in order to bound $\Big\|P_f^\xi|\mathscr{C}_{\bar A_T^n(x)}-P_f^{n,\bar\xi}|\mathscr{C}_{\bar A_T^n(x)}\Big\|_{\mathrm{TV}}$ and hence $\Delta(\mathscr{P}_\xi^{\bar A_T^n(x)},\mathscr{P}_{\bar\xi}^{n,\bar A_T^n(x)})$. More precisely, let h_f be the Hellinger process of order 1/2 between the measures $P_f^\xi|\mathscr{C}_{\bar A_T^n(x)}$ and $P_f^{n,\bar\xi}|\mathscr{C}_{\bar A_T^n(x)}$, that is, (see Jacod and Shiryaev, Jacod, Shiryaev (1987), page 239)

$$h_f(t)(x)=\frac{1}{8\varepsilon^2}\int_0^t\bigg(\frac{f(x_s)}{\sigma^2(x_s)}-\bar g_n(s,x)\bigg)^2ds.$$

Then:

$$\Big\|P_f^\xi|_{\mathscr{C}_{\bar A_T^n(x)}}-P_f^{n,\bar\xi}|_{\mathscr{C}_{\bar A_T^n(x)}}\Big\|_{\mathrm{TV}}\le 4\sqrt{\mathbb{E}_{P_f^\xi}h_f(\bar A_T^n(x))(x)},$$

as in Jacod, Shiryaev (1987), 4b, Theorem 4.21, page 279. Hence we conclude thanks to Lemma 6.4.8. □

We now prove that $\Delta(\mathscr{P}_\xi^{A_T^n(x)}, \mathscr{P}_\xi^{S_T^n(x)}) \to 0$ and $\Delta(\mathscr{P}_\xi^{S_T^n(x)}, \mathscr{P}_\xi^{\bar{A}_T^n(x)}) \to 0$ as $n \to \infty$. Again, we start with a lemma:

Lemma 6.4.10. *Under the hypotheses of Proposition 6.4.3 and with the same notation as in Steps 1 and 2, we have*

$$\mathbb{E}_{P_f^\xi}|A_T(x) - \bar{A}_T^n(x)| = O\Big(\frac{1}{n} + \varepsilon\Big), \tag{6.15}$$

uniformly over $\mathscr{F}$.

Proof. The crucial point in proving (6.15) is to use the convergence of diffusion processes with small variance to some deterministic solution. To that aim, let us introduce the following ODEs:

$$\frac{dz_t}{dt} = f(z_t), \quad \frac{d\bar{z}_t}{dt} = \frac{f(\bar{z}_t)}{\sigma^2(\bar{z}_t)}\bar{\sigma}_n^2(t, \bar{z}), \quad z_0 = w = \bar{z}_0.$$

By means of Property 6.4.2, the Lipschitz character of $\sigma^2(\cdot)$ and the linear growth of f, we get,

$$\begin{aligned}
\mathbb{E}_{P_f^\xi}|A_T(x) - \bar{A}_T^n(x)| &= \mathbb{E}_{P_f^\xi}\Big|\int_0^T \big(\sigma^2(x_{A_t(x)}) - \bar{\sigma}_n^2(t, x_{\bar{A}^n_\cdot(x)})\big)dt\Big| \\
&\leq 2\sigma_1 M \mathbb{E}_{P_f^\xi}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} |x_{A_t(x)} - x_{\bar{A}_{t_i}^n(x)}|dt \\
&\leq 2\sigma_1 M \mathbb{E}_{P_f^\xi}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} \big(|x_{A_t(x)} - z_t| + |z_t - \bar{z}_{t_i}| + |\bar{z}_{t_i} - x_{\bar{A}_{t_i}^n(x)}|\big)dt.
\end{aligned}$$

For all $t \in [t_i, t_{i+1}]$, we shall analyze the terms $I = \mathbb{E}_{P_f^\xi}|x_{A_t(x)} - z_t|$, $II = |z_t - \bar{z}_{t_i}|$ and $III = \mathbb{E}_{P_f^\xi}|\bar{z}_{t_i} - x_{\bar{A}_{t_i}^n(x)}|$, separately.

- Term I: By means of Lemma 6.4.6 and some standard calculations one can write

$$\begin{aligned}
\mathbb{E}_{P_f^\xi}|x_{A_t(x)} - z_t| = \mathbb{E}_{P_f^y}|x_t - z_t| &= \mathbb{E}_{P_f^y}\Big|\int_0^t \big(f(x_s) - f(z_s)\big)ds + \varepsilon\int_0^t \sigma(x_s)dW_s\Big| \\
&\leq M\mathbb{E}_{P_f^y}\int_0^t |x_s - z_s|ds + \varepsilon\sigma_1\sqrt{t};
\end{aligned}$$

  hence, an application of the Gronwall lemma yields $\mathbb{E}_{P_f^\xi}|x_{A_t(x)} - z_t| \leq \varepsilon\sqrt{t}\exp(MT)$.

- Term II: By the triangular inequality it is enough to bound $|z_t - z_{t_i}|$ and $|z_{t_i} - \bar{z}_{t_i}|$, separately. It is easy to see that $|z_s - z_{t_j}|$ is a $O(n^{-1})$ as well as $|\bar{z}_s - \bar{z}_{t_j}|$ for all $s \in [t_j, t_{j+1}]$, $j = 0, \dots, n-1$. Moreover, observe that there exists a constant C, independent of f, such that $\left|f(x) - \frac{f(y)}{\sigma^2(y)}\sigma^2(z)\right| \leq C(|x-y| + (1+|y|)|y-z|)$. We get:

$$\begin{aligned}
|z_{t_i} - \bar{z}_{t_i}| &= \left| \int_0^{t_i} \Big(f(z_s) - \frac{f(\bar{z}_s)}{\sigma^2(\bar{z}_s)} \bar{\sigma}_n^2(s, \bar{z}) \Big) ds \right| \\
&= \left| \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \Big(f(z_s) - \frac{f(\bar{z}_s)}{\sigma^2(\bar{z}_s)} \sigma^2(\bar{z}_{t_j}) \Big) ds \right| \\
&\leq C \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \big(|z_s - \bar{z}_s| + (1 + |\bar{z}_s|)|\bar{z}_s - \bar{z}_{t_j}| \big) ds \\
&\leq C \int_0^{t_i} |z_s - \bar{z}_s| ds + \frac{C' t_i}{n},
\end{aligned}$$

for some constant C', independent of f. Therefore, applying the Gronwall lemma one obtains

$$|z_{t_i} - \bar{z}_{t_i}| \leq \frac{C' t_i}{n} e^{C t_i}$$

that allows us to conclude $|z_t - \bar{z}_{t_i}| = O(n^{-1})$.

- Term III: By means of Lemma 6.4.6 we know that $\mathbb{E}_{P_f^{\xi}} |\bar{z}_{t_i} - x_{\bar{A}^n_{t_i}(x)}| = \mathbb{E}_{P_f^{\bar{\zeta}}} |\bar{z}_{t_i} - x_{t_i}|$.

$$\begin{aligned}
&\mathbb{E}_{P_f^{\bar{\zeta}}} |\bar{z}_{t_i} - x_{t_i}| = \mathbb{E}_{P_f^{\bar{\zeta}}} \left| \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \Big[\Big(\frac{f(\bar{z}_s)}{\sigma^2(\bar{z}_s)} \sigma^2(\bar{z}_{t_j}) - \frac{f(x_s)}{\sigma^2(x_s)} \sigma^2(x_{t_j}) \Big) ds + \varepsilon \sigma(x_{t_j}) dW_s \Big] \right| \\
&\leq \mathbb{E}_{P_f^{\bar{\zeta}}} \left| \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \Big(\frac{f(\bar{z}_s)}{\sigma^2(\bar{z}_s)} \sigma^2(\bar{z}_{t_j}) - \frac{f(x_s)}{\sigma^2(x_s)} \sigma^2(x_{t_j}) \Big) ds \right| + \varepsilon \sigma_1 \sqrt{t_i} \\
&\leq \mathbb{E}_{P_f^{\bar{\zeta}}} \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \bigg(\frac{M\sigma_1^2}{\sigma_0^2} |\bar{z}_s - x_s| + \frac{2\sigma_1 M}{\sigma_0^4} (1 + |\bar{z}_s|)(|\bar{z}_{t_j} - \bar{z}_s| + |x_s - x_{t_j}|) \bigg) ds + \varepsilon \sigma_1 \sqrt{t_i} \\
&\leq \mathbb{E}_{P_f^{\bar{\zeta}}} \int_0^{t_i} \frac{M\sigma_1^2}{\sigma_0^2} |\bar{z}_s - x_s| ds + C n^{-1} t_i + \varepsilon \sigma_1 \sqrt{t_i},
\end{aligned}$$

for some constant C independent of f. An application of the Gronwall lemma gives

$$\mathbb{E}_{P_f^{\bar{\zeta}}} |\bar{z}_{t_i} - x_{t_i}| \leq \big(C n^{-1} t_i + \varepsilon \sigma_1 \sqrt{t_i} \big) \exp\Big(\frac{M\sigma_1^2}{\sigma_0^2} t_i \Big).$$

Putting all the pieces together we obtain $\mathbb{E}_{P_f^\xi}|A_T(x)-\bar A_T^n(x)| = O\Big(\frac{1}{n}+\varepsilon\Big)$. $\square$

Proposition 6.4.11. *Under the same hypotheses of Proposition 6.4.3, we have that* $\Delta(\mathscr{P}_\xi^{A_T(x)},\mathscr{P}_\xi^{S_T^n(x)})\to 0$ *and* $\Delta(\mathscr{P}_\xi^{S_T^n(x)},\mathscr{P}_\xi^{\bar A_T^n(x)})\to 0$ *as* $n\to\infty$.

Proof. We shall prove only the first statement, the proof of the second one being identical. Since $\mathscr{C}_{S_T^n(x)}\subset\mathscr{C}_{A_T(x)}$, it is clear that $\delta(\mathscr{P}_\xi^{A_T(x)},\mathscr{P}_\xi^{S_T^n(x)})=0$. To control $\delta(\mathscr{P}_\xi^{S_T^n(x)},\mathscr{P}_\xi^{A_T(x)})$ we will introduce the following Markov kernel K^n:

$$K^n(\omega,A):=\mathbb{E}_{P_0^\xi}\big(\mathbb{I}_A|_{\mathscr{C}_{S_T^n(x)}}\big)(\omega),\quad \forall A\in\mathscr{C}_{A_T(x)},\omega\in C,$$

where P_0^ξ is defined as P_f^ξ with $f\equiv 0$. Remark that the Markov kernel K^n thus constructed coincides with the Markov kernel N defined in Genon-Catalot, Laredo (2014), Proposition 6.2, when $\varepsilon\equiv 1$. Making the same computations as in the cited proposition, we obtain that

$$\begin{aligned}\big\|K^nP_f^\xi|_{\mathscr{C}_{S_T^n(x)}}-P_f^\xi|_{\mathscr{C}_{A_T(x)}}\big\|_{TV}&\le\frac{1}{2}\sqrt{\mathbb{E}_{P_f^\xi|\mathscr{C}_{A_T(x)}}\int_{S_T^n(x)}^{A_T(x)}\frac{f^2(x_r)}{\sigma^4(x_r)}dr}\\&\le\frac{M}{\sqrt{2}\sigma_0^2}\sqrt{\mathbb{E}_{P_f^\xi|\mathscr{C}_{A_T(x)}}\bigg(|A_T(x)-\bar A_T^n(x)|+\int_0^{|A_T(x)-\bar A_T^n(x)|}x_r^2dr\bigg)}\\&=O\Big(\big(\mathbb{E}_{P_f^\xi|\mathscr{C}_{A_T(x)}}|A_T(x)-\bar A_T^n(x)|\big)^{1/4}\Big).\end{aligned}$$

We then conclude that $\Delta(\mathscr{P}_\xi^{A_T(x)},\mathscr{P}_\xi^{S_T^n(x)})\to 0$ by means of Lemma 6.4.10. $\square$

Step 4. Using Steps 1–3 and the triangular inequality, one can find that

$$\Delta(\mathscr{P}_y^T,\mathscr{P}_{\bar y}^{n,T})=O\Big(\frac{1}{\varepsilon n}+(n^{-1}+\varepsilon)^{1/4}\Big).$$

Hence, to conclude the proof of Theorem 6.3.1, we only need to show the following proposition:

Proposition 6.4.12. *Under the same hypotheses of Proposition 6.4.3,* $\Delta(\mathscr{P}_{\bar y}^{n,T},\mathscr{Q}_Z^n)=0$, *for all* n.

Proof. Note that, by using the Girsanov theorem, we can show that the measure $P_f^{n,\bar y}|\mathscr{C}_T$ is absolutely continuous with respect to $P_0^{n,\bar y}$ and the density is given by

$$\frac{dP_f^{n,\bar y}}{dP_0^{n,\bar y}}|_{\mathscr{C}_T}(\omega)=\exp\bigg(\sum_{i=0}^{n-1}\Big(\frac{f(\omega_{t_i})}{\varepsilon^2\sigma^2(\omega_{t_i})}(\omega_{t_{i+1}}-\omega_{t_i})-\frac{f^2(\omega_{t_i})}{2n\varepsilon^2\sigma^2(\omega_{t_i})}\Big)\bigg).$$

Hence, by means of the Fisher's factorization theorem, we can deduce that the application $S:\omega\to(\omega_{t_1},\dots,\omega_{t_n})$ is a sufficient statistic for the family of probability measures $\{P_f^{n,\bar y}|_{\mathscr{C}_T}; f\in\mathscr{F}\}$. We complete the proof remarking that the distribution of $(x_{t_1},\dots,x_{t_n})$ under $P_f^{n,\bar y}$ is the same as the one of $(Z_1,\dots,Z_n)$ under $\mathbb{P}$ and finally invoking the following property of the Le Cam distance (see Le Cam (1986)):

Let $\mathscr{P}_i=(\mathscr{X}_i,\mathscr{A}_i,\{P_{i,\theta},\theta\in\Theta\})$, $i=1,2$, be two statistical models and let $(\mathscr{X}_1,\mathscr{A}_1)$ be a Polish space. Let $S:\mathscr{X}_1\to\mathscr{X}_2$ be a sufficient statistics such that the distribution of S under $P_{1,\theta}$ is equal to $P_{2,\theta}$. Then $\Delta(\mathscr{P}_1,\mathscr{P}_2)=0$. □

6.4.3 Proof of Theorem 6.3.2

We will proceed in three Steps.

Step 1. Let us consider the application $F:\mathbb{R}\to\mathbb{R}$ defined as $F(x)=\int_0^x\frac{1}{\varepsilon\sigma(u)}du$. Remark that F is well defined and one to one. Using the Itô formula, we have that

$$F(y_t)=F(w)+\int_0^t\Big(\frac{f(y_s)}{\varepsilon\sigma(y_s)}-\frac{\varepsilon\sigma'(y_s)}{2}\Big)ds+W_t.$$

Thus, if we set $\mu_t:=F(y_t)$, the new process (μ_t) satisfies the following SDE:

$$\mu_0=F(w);\quad d\mu_t=\Big(\frac{f(F^{-1}(\mu_t))}{\varepsilon\sigma(F^{-1}(\mu_t))}-\frac{\varepsilon\sigma'(F^{-1}(\mu_t))}{2}\Big)dt+dW_t,\quad t\in[0,T]. \tag{6.16}$$

Observe that, thanks to hypotheses (H1), (H2) and the Lipschitz condition on $\frac{f}{\sigma}(\cdot)$, the drift function $b(x):=\frac{f(F^{-1}(x))}{\varepsilon\sigma(F^{-1}(x))}-\frac{\varepsilon\sigma'(F^{-1}(x))}{2}$ is such that $|b(0)|\le\frac{M}{\varepsilon\sigma_0}+\frac{\varepsilon M}{2}$ and it is also Lipschitz:

$$\begin{aligned}|b(x)-b(y)|&=\Big|\Big(\frac{f(F^{-1}(x))}{\varepsilon\sigma(F^{-1}(x))}-\frac{\varepsilon\sigma'(F^{-1}(x))}{2}\Big)-\Big(\frac{f(F^{-1}(y))}{\varepsilon\sigma(F^{-1}(y))}-\frac{\varepsilon\sigma'(F^{-1}(y))}{2}\Big)\Big|\\&\le\frac{1}{\varepsilon}\Big|\frac{f(F^{-1}(x))}{\sigma(F^{-1}(x))}-\frac{f(F^{-1}(y))}{\sigma(F^{-1}(y))}\Big|+\varepsilon\Big|\frac{\sigma'(F^{-1}(x))}{2}-\frac{\sigma'(F^{-1}(y))}{2}\Big|\\&\le\frac{L}{\varepsilon}|F^{-1}(x)-F^{-1}(y)|+\frac{M\varepsilon}{2}|F^{-1}(x)-F^{-1}(y)|\\&=\Big|\int_x^y\varepsilon\sigma(F(u))du\Big|\Big(\frac{L}{\varepsilon}+\frac{M\varepsilon}{2}\Big)\le\Big(\frac{L}{\varepsilon}+\frac{M\varepsilon}{2}\Big)\sigma_1\varepsilon|x-y|;\end{aligned}$$

In particular the existence and the uniqueness of a strong solution μ for the SDE (6.16) are guaranteed. Let us denote by P_f^μ (resp. $Q_f^{n,\mu}$) the law of μ (resp. $(\mu_{t_1},\dots,\mu_{t_n})$) and introduce the statistical models

$$\mathscr{P}_\mu^T=\big(C,\mathscr{C}_T,(P_f^\mu,f\in\mathscr{F})\big),\quad\mathscr{Q}_\mu^n=\big(\mathbb{R}^n,\mathscr{B}(\mathbb{R}^n),(Q_f^{n,\mu},f\in\mathscr{F})\big).$$

By construction, $\mathscr{P}_\mu^T$ (resp. $\mathscr{Q}_\mu^n$) is the image experiment of $\mathscr{P}_y^T$ (resp. $\mathscr{Q}_y^n$) by F. Thus we have:

Proposition 6.4.13. *Under the hypotheses of Theorem 6.3.2, the statistical models $\mathscr{P}_y^T$ (resp. $\mathscr{Q}_y^n$) and $\mathscr{P}_\mu^T$ (resp. $\mathscr{Q}_\mu^n$) are equivalent.*

Step 2. Using the same notations as above, define a new drift function $\bar{b}_n$:

$$\bar{b}_n(t,\omega) = \sum_{i=0}^{n-1} b(\omega_{t_i})\mathbb{I}_{(t_i,t_{i+1}]}(t), \quad \forall \omega \in C,\ t \in [0,T]$$

and consider the diffusion process $(\bar{\mu}_t)$ on $(C, \mathscr{C}_T)$ having drift function given by $\bar{b}_n$ and diffusion coefficient equal to 1, i.e.

$$\bar{\mu}_0 = F(w); \quad d\bar{\mu}_t = \bar{b}_n(t,\bar{\mu})dt + dW_t, \quad t \in [0,T]. \tag{6.17}$$

Denote by $P_f^{n,\bar{\mu}}$ the law of the solution of (6.17) and introduce the corresponding statistical model:

$$\mathscr{P}_{\bar{\mu}}^{n,T} = \big(C, \mathscr{C}_T, (P_f^{n,\bar{\mu}}, f \in \mathscr{F})\big).$$

Proposition 6.4.14. *Under the hypotheses of Theorem 6.3.2, the statistical models $\mathscr{P}_\mu^{n,T}$ and $\mathscr{P}_{\bar{\mu}}^{n,T}$ are asymptotically equivalent as n goes to infinity and T is fixed.*

Proof. One can use the same arguments as in the proof of Proposition 6.4.9 obtaining a first bound given by

$$\Delta(\mathscr{P}_\mu^T, \mathscr{P}_{\bar{\mu}}^{n,T}) \le \sup_{f \in \mathscr{F}} \|P_f^\mu - P_f^{n,\bar{\mu}}\|_{TV} \le 4 \sup_{f \in \mathscr{F}} \sqrt{\mathbb{E}_{P_f^\mu} \frac{1}{8} \int_0^T (b(x_s) - \bar{b}_n(s,x))^2 ds}.$$

Now, thanks to the $\tilde{L}$-Lipschitz character of b, one can write $\int_0^T (b(x_s) - \bar{b}_n(s,x))^2 ds \le \tilde{L} \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} (x_s - x_{t_i})^2 ds$ so that the usual computations yield $\Delta(\mathscr{P}_\mu^T, \mathscr{P}_{\bar{\mu}}^{n,T}) = O(n^{-1})$. □

Step 3. Consider now the statistical model associated with the discrete observations $(\bar{\mu}_{t_1}, \dots, \bar{\mu}_{t_n})$:

$$\mathscr{Q}_{\bar{\mu}}^n = \big(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), (Q_f^{n,\bar{\mu}}, f \in \mathscr{F})\big),$$

where $Q_f^{n,\bar{\mu}}$ denotes the law of the vector $(\bar{\mu}_{t_1}, \dots, \bar{\mu}_{t_n})$.

Proposition 6.4.15. *Under the hypotheses of Theorem 6.3.2, we have*

$$\Delta(\mathscr{P}_{\bar{\mu}}^{n,T}, \mathscr{Q}_{\bar{\mu}}^n) = 0, \quad \Delta(\mathscr{Q}_{\bar{\mu}}^n, \mathscr{Q}_\mu^n) = O\big(n^{-1}\big),\ \forall n.$$

Proof. The first equivalence can be proved by means of a sufficient statistic as in the proof of Proposition 6.4.12; the second one follows directly from Step 2 since $\|Q_f^{n,\mu} - Q_f^{n,\bar{\mu}}\|_{TV} \le \|P_f^\mu - P_f^{n,\bar{\mu}}\|_{TV}$ as we are only restricting to a smaller σ-algebra. □

Background on Le Cam's theory

Asymptotic equivalence in the sense of Le Cam

A *statistical model* is a triplet $\mathscr{P}_j = (\mathscr{X}_j, \mathscr{A}_j, \{P_{j,\theta}; \theta \in \Theta\})$ where $\{P_{j,\theta}; \theta \in \Theta\}$ is a family of probability distributions all defined on the same σ-field $\mathscr{A}_j$ over the *sample space* $\mathscr{X}_j$ and Θ is the *parameter space*. The *deficiency* $\delta(\mathscr{P}_1, \mathscr{P}_2)$ of $\mathscr{P}_1$ with respect to $\mathscr{P}_2$ quantifies "how much information we lose" by using $\mathscr{P}_1$ instead of $\mathscr{P}_2$ and is defined as $\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_K \sup_{\theta\in\Theta} ||KP_{1,\theta} - P_{2,\theta}||_{TV}$, where TV stands for "total variation" and the infimum is taken over all "transitions" K (see Le Cam (1986), page 18). In our setting, however, the general notion of "transitions" can be replaced with the notion of Markov kernels. Indeed, when the model $\mathscr{P}_1$ is dominated and the sample space $(\mathscr{X}_2, \mathscr{A}_2)$ of the experiment $\mathscr{P}_2$ is a Polish space, the infimum appearing on the definition of the deficiency δ can be taken over all Markov kernels K on $\mathscr{X}_1 \times \mathscr{A}_2$ (see Nussbaum (1996), Proposition 10.2), i.e.

$$\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_K \sup_{\theta\in\Theta} \sup_{A\in\mathscr{A}_2} \left| \int_{\mathscr{X}_1} K(x, A) P_{1,\theta}(dx) - P_{2,\theta}(A) \right|. \tag{6.18}$$

The experiment $KP_{1,\theta} = (\mathscr{X}_1, \mathscr{A}_1, \{KP_{1,\theta}\}_{\theta\in\Theta})$ is called a *randomization* of $\mathscr{P}_1$ by the kernel K. If the kernel is deterministic, i.e. for $T : (\mathscr{X}_1, \mathscr{A}_1) \to (\mathscr{X}_2, \mathscr{A}_2)$ a random variable, $T(x, A) := \mathbb{I}_A(T(x))$, the experiment $T\mathscr{P}_1$ is called the *image experiment by the random variable* T. Closely associated with the notion of deficiency is the so called Δ-distance, i.e. the pseudo metric defined by:

$$\Delta(\mathscr{P}_1, \mathscr{P}_2) := \max(\delta(\mathscr{P}_1, \mathscr{P}_2), \delta(\mathscr{P}_2, \mathscr{P}_1)).$$

The sufficiency of a statistic can be expressed in terms of the Δ-distance. More precisely, the following holds (see Genon-Catalot, Laredo (2014), Proposition 8.1, page 23). *Let $T : (\mathscr{X}_1, \mathscr{A}_1) \to (\mathscr{X}_2, \mathscr{A}_2)$ be a random variable. The statistic T is sufficient for $\mathscr{P}_1$ if and only if $\Delta(\mathscr{P}_1, T\mathscr{P}_1) = 0$.*

Also, remark that thanks to (6.18), if the statistical models $\mathscr{P}_1 = (\mathscr{X}, \mathscr{A}_1, \{P_\theta; \theta \in \Theta\})$ and $\mathscr{P}_2 = (\mathscr{X}, \mathscr{A}_2, \{P_\theta; \theta \in \Theta\})$ with $\mathscr{A}_2 \subset \mathscr{A}_1$, then $\delta(\mathscr{P}_1, \mathscr{P}_2) = 0$.

Two sequences of statistical models $(\mathscr{P}_1^n)_{n\in\mathbb{N}}$ and $(\mathscr{P}_2^n)_{n\in\mathbb{N}}$ are called *asymptotically equivalent* if $\Delta(\mathscr{P}_1^n, \mathscr{P}_2^n)$ tends to zero as n goes to infinity. Similarly, the statistic T^n is *asymptotically sufficient* for $\mathscr{P}_1^n$ if $\Delta(\mathscr{P}_1^n, T^n\mathscr{P}_1^n)$ tends to zero as n goes to infinity.

Chapter 7

L_1-distance for additive processes with time-homogeneous Lévy measures

Résumé Au cours du Chapitre 7 nous proposons une majoration de la distance L_1 entre deux processus additifs avec une caractéristique locale $(f_j(\cdot), \sigma^2(\cdot), \nu_j)$, $j = 1, 2$. Les cas $\sigma = 0$ et $\sigma(\cdot) > 0$ sont traités. Les mesures de Lévy ν_1 et ν_2 peuvent être éventuellement à variation infinie. Le Chapitre 7 est basé sur un article publié dans *Electronic Communications in Probability.* Il s'agit d'un travail commun avec Pierre Étoré.

Mot clés: distance L_1, processus additifs, processus de Lévy.

Abstract We give an explicit bound for the L_1-distance between two additive processes of local characteristics $(f_j(\cdot), \sigma^2(\cdot), \nu_j)$, $j = 1, 2$. The cases $\sigma = 0$ and $\sigma(\cdot) > 0$ are both treated. We allow ν_1 and ν_2 to be time-homogeneous Lévy measures, possibly with infinite variation. Chapter 7 is based on a paper published in *Electronic Communications in Probability.* It is a joint work with Pierre Étoré.

Keywords: L_1-distance, additive processes, Lévy processes.

7.1 Introduction and main result

In this note we give an upper bound for the L_1-distance between the laws induced on the Skorokhod space by two additive processes observed until time $T > 0$. By the L_1-distance between two σ-finite measures μ_1 and μ_2 on $(E,\mathscr{E})$ such that μ_1 is absolutely continuous with respect to μ_2 we mean

$$L_1(\mu_1,\mu_2)=2\sup_{A\in\mathscr{E}}\big|\mu_1(A)-\mu_2(A)\big|=\int_E\bigg|\frac{d\mu_1}{d\mu_2}-1\bigg|d\mu_2.$$

Note that, with our definitions, the L_1-distance is twice the so called total variation distance.

Giving bounds for the L_1-distance is a classical problem, which, in the last decades, has been reinterpreted in more modern terms via Stein's method (see, e.g., Nourdin, Peccati (2009); Peccati (2011); Ross (2011)). This kind of problems arises in several fields such us Bayesian statistics, convergence rates of Markov chains or Monte Carlo algorithms (see Gibbs, Su (2002), Section 4 and the references therein). However, to the best of our knowledge, results bounding the L_1-distance between laws on the Skorokhod space are much less abundant. In this setting other kinds of distances have been privileged such as the Wasserstein-Kantorovich-Rubinstein metric (see Gairing et al. (2013)). More relevant to our purposes is a result due to Memin, Shiryayev (1985) computing the Hellinger distance between the laws of any two processes with independent increments. In particular this gives a bound for the L_1-distance between additive processes. In order to state their result let us fix some notation.

Let $\{x_t\}$ be the canonical process on the Skorokhod space $(D,\mathscr{D})$ and denote by $P^{(f,\sigma^2,\nu)}$ the law induced on $(D,\mathscr{D})$ by an additive process having local characteristics $(f(\cdot),\sigma^2(\cdot),\nu)$. We will denote such a process by $\big(\{x_t\},P^{(f,\sigma^2,\nu)}\big)$ and we will write $P_T^{(f,\sigma^2,\nu)}$ for the restriction of $P^{(f,\sigma^2,\nu)}$ to the σ-algebra generated by $\{x_s:0\leq s\leq T\}$ (see Section 7.2 for the precise definitions). Our purpose is to bound $L_1\big(P_T^{(f_1,\sigma_1^2,\nu_1)},P_T^{(f_2,\sigma_2^2,\nu_2)}\big)$. From now on we will assume that $\sigma_1^2(\cdot)=\sigma_2^2(\cdot)=\sigma^2(\cdot)$, otherwise this distance is 2 (see, e.g., Jacod, Shiryaev (2003); Newman (1972)). We also need to define the following quantities:

$$\gamma^{\nu_j}=\int_{|y|\leq 1}y\nu_j(dy),\quad j=1,2;\quad \xi^2=\int_0^T\frac{(f_2(r)-f_1(r)-(\gamma^{\nu_2}-\gamma^{\nu_1}))^2}{\sigma^2(r)}dr.$$

Theorem 7.1.1 (Memin and Shiryayev)**.** *Let $\big(\{x_t\},P^{(f_1,\sigma^2,\nu_1)}\big)$ and $\big(\{x_t\},P^{(f_2,\sigma^2,\nu_2)}\big)$ be two additive processes with ν_1 and ν_2 Lévy measures such that ν_1 is absolutely continuous with respect to ν_2 and satisfying:*

$$H^2(\nu_1,\nu_2):=\int_{\mathbb{R}}\bigg(\sqrt{\frac{d\nu_1}{d\nu_2}(y)}-1\bigg)^2\nu_2(dy)<\infty. \tag{7.1}$$

The following upper bounds hold, for any $0<T<\infty$: If $\sigma^2>0$ then

$$L_1\Big(P_T^{(f_1,\sigma^2,\nu_1)},P_T^{(f_2,\sigma^2,\nu_2)}\Big)\leq\sqrt{8\Big(1-\exp\Big(-\frac{\xi^2}{8}-\frac{T}{2}H^2(\nu_1,\nu_2)\Big)\Big)}.$$

If $\sigma^2 = 0$ *and* $f_1 - f_2 \equiv \gamma^{\nu_1} - \gamma^{\nu_2}$, *then*

$$L_1\Big(P_T^{(f_1,0,\nu_1)}, P_T^{(f_2,0,\nu_2)}\Big) \leq \sqrt{8\Big(1 - \exp\Big(-\frac{T}{2}H^2(\nu_1,\nu_2)\Big)\Big)}.$$

Observe that (7.1) implies $\gamma^{\nu_j} < \infty$, $j = 1, 2$. When $\sigma^2 = 0$ it follows from Theorem 7.1.2 that, for example, $L_1\Big(P_T^{(\gamma^{\nu_1},0,\nu_1)}, P_T^{(\gamma^{\nu_2},0,\nu_2)}\Big) \leq 2\sqrt{TL_1(\nu_1,\nu_2)}$.

The proof of Theorem 7.1.1, however, makes heavy use of general theory of semimartingales. This note originated from the research for a proof based only on classical results for Lévy processes, Esscher-type transformations and the Cameron-Martin formula. It turned out that this method, when applicable, gives sharper bound on the L_1-distance. More precisely, our main result is as follows.

Theorem 7.1.2. *Let* $\big(\{x_t\}, P^{(f_1,\sigma^2,\nu_1)}\big)$ *and* $\big(\{x_t\}, P^{(f_2,\sigma^2,\nu_2)}\big)$ *be two additive processes with* ν_1 *and* ν_2 *Lévy measures such that* ν_1 *is absolutely continuous with respect to* ν_2 *and satisfying:*

$$L_1(\nu_1,\nu_2) < \infty.$$

Then, the following upper bounds hold, for any $0 < T < \infty$.

If $\sigma^2 > 0$ *then*

$$L_1\Big(P_T^{(f_1,\sigma^2,\nu_1)}, P_T^{(f_2,\sigma^2,\nu_2)}\Big) \leq 2\sinh\Big(TL_1(\nu_1,\nu_2)\Big) + 2\Big[1 - 2\phi\Big(-\frac{\xi}{2}\Big)\Big].$$

If $\sigma^2 = 0$ *and* $f_1 - f_2 \equiv \gamma^{\nu_1} - \gamma^{\nu_2}$, *then*

$$L_1\Big(P_T^{(f_1,0,\nu_1)}, P_T^{(f_2,0,\nu_2)}\Big) \leq 2\sinh\Big(TL_1(\nu_1,\nu_2)\Big).$$

Remark that, in the case $\nu_1 = \nu_2 = 0$, i.e. where there are no jumps, the upper bound in Theorem 7.1.2 is achieved. Indeed, an explicit formula for the L_1-distance between Gaussian processes is well known. Denoting by ϕ the cumulative distribution function of a normal random variable $\mathscr{N}(0,1)$, we have, for any $0 < T < \infty$:

$$L_1\big(P_T^{(f_1,\sigma^2,0)}, P_T^{(f_2,\sigma^2,0)}\big) = 2\Big(1 - 2\phi\Big(-\frac{1}{2}\sqrt{\int_0^T \frac{(f_1(t)-f_2(t))^2}{\sigma_1^2(t)}dt}\Big)\Big)$$

whenever the right-hand side term makes sense (see, e.g., Brown, Low (1996)).

The reason for our interest in the L_1-distance lies in the fact that it is a fundamental tool in the Le Cam theory of comparison of statistical models (Le Cam (1986); Le Cam, Yang (2000)). More precisely, the presented result will be needed in a forthcoming paper by the second author, establishing an equivalence result, in the Le Cam sense, for additive processes. Similar estimations appear in many other results concerning the Le Cam distance. See for example Brown, Low (1996); Carter (2002); Nussbaum (1996), where the L_1-distance between Gaussian processes is computed or Dalalyan, Reiß (2006); Genon-Catalot, Laredo (2014); Milstein, Nussbaum (1998) concerning diffusion processes without jumps. In recent years, however, there is a growing interest in models with jumps due to their numerous applications in econometrics, insurance theory or financial modelling. Because of that, it is useful to dispose of simple formulas for estimating distances between such processes.

Theorem 7.1.2 is proved in Section 7.3. In Section 7.2 we collect some preliminary results about additive processes that will play a role in the proof. Before that, we give some examples of situations where our result can be applied. The choice of these examples are inspired by the models exhibited in Cont, Tankov (2004).

Example 7.1.3. (L_1-distance between compound Poisson processes)

Let $\{X_t^1\}$ and $\{X_t^2\}$ be two compound Poisson processes on $[0,T]$ with intensities $\lambda_j > 0$, $j = 1,2$ and jump size distributions G_j; i.e. $\{X_t^j\}$ is a Lévy process of characteristic triplet $\big(\lambda_j \int_{|y|\leq 1} yG_j(dy), 0, \lambda_j G_j\big)$. Furthermore, let A be a subset of $\mathbb{R}$ and suppose that G_j is equivalent to the Lebesgue measure restricted to A. Denote by g_j the density $\frac{dG_j}{d\mathrm{Leb}_{|A}}$; then, an application of Theorem 7.1.2 yields:

$$L_1\big(X^1, X^2\big) \leq 2\sinh\Big(T\int_A |\lambda_1 g_1(y) - \lambda_2 g_2(y)|dy\Big).$$

Example 7.1.4. (L_1-distance between additive processes of jump-diffusion type)

An additive process of jump-diffusion type on $[0,T]$ has the following form:

$$X_t = \int_0^t f(r)dr + \int_0^t \sigma(r)dW_r + \sum_{i=1}^{N_t} Y_i, \quad t \in [0,T],$$

where $\{W_t\}$ is a standard Brownian motion, $\{N_t\}$ is the Poisson process counting the jumps of $\{X_t\}$, and Y_i are jumps sizes (i.i.d. random variables). Consider now the additive processes of jump-diffusion type $\{X_t^j\}$ having local characteristics given by $(f_j(\cdot) + \lambda_j \int_{|y|\leq 1} yG_j(dy), \sigma^2(\cdot), \lambda_j G_j)$, $j = 1,2$ and suppose again that G_j is equivalent to the Lebesgue measure restricted to some $A \subseteq \mathbb{R}$. Letting g_j denote the density of G_j as above, we have:

$$L_1\big(X^1, X^2\big) \leq 2\sinh\Big(T\int_A |\lambda_1 g_1(y) - \lambda_2 g_2(y)|dy\Big) + 2\Big(1 - 2\phi\Big(-\sqrt{\int_0^T \frac{(f_1(t) - f_2(t))^2}{4\sigma^2(t)}dt}\Big)\Big).$$

Example 7.1.5. (L_1-distance between tempered stable processes)

Let $\{X_t^1\}$ and $\{X_t^2\}$ be two tempered stable processes, i.e. Lévy processes on $\mathbb{R}$ with no gaussian component and such that their Lévy measures ν_j have densities of the form

$$\frac{d\nu_j}{d\mathrm{Leb}}(y) = \frac{C_-}{|y|^{1+\alpha}} e^{-\lambda_-^j |y|} \mathbb{I}_{y<0} + \frac{C_+}{y^{1+\alpha}} e^{-\lambda_+^j y} \mathbb{I}_{y>0}, \quad j = 1, 2,$$

for some parameters $C_\pm > 0$, $\lambda_\pm^j > 0$ and $\alpha < 2$. Then the hypothesis (7.1) is satisfied and Theorem 7.1.2 bounds the L_1-distance by:

$$2\sinh\left(T\left[C_+ \int_0^\infty \left|\frac{e^{-\lambda_+^1 y} - e^{-\lambda_+^2 y}}{y^{1+\alpha}}\right| dy + C_- \int_{-\infty}^0 \left|\frac{e^{-\lambda_-^1 |y|} - e^{-\lambda_-^2 |y|}}{|y|^{1+\alpha}}\right| dy\right]\right).$$

7.2 Preliminary results

7.2.1 Additive processes

Definition 7.2.1. A stochastic process $\{X_t\} = \{X_t : t \geq 0\}$ on $\mathbb{R}$ defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ is an *additive process* if the following conditions are satisfied.

1. $X_0 = 0$ $\mathbb{P}$-a.s.
2. For any choice of $n \geq 1$ and $0 \leq t_0 < t_1 < \ldots < t_n$, random variables X_{t_0}, $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$are independent.
3. There is $\Omega_0 \in \mathscr{A}$ with $\mathbb{P}(\Omega_0) = 1$ such that, for every $\omega \in \Omega_0$, $X_t(\omega)$ is right-continuous in $t \geq 0$ and has left limits in $t > 0$.
4. It is stochastically continuous.

Thanks to the *Lévy-Khintchine formula*, the characteristic function of any additive process $\{X_t\}$ can be expressed, for all u in $\mathbb{R}$, as:

$$\mathbb{E}\big[e^{iuX_t}\big] = \exp\Big(iu\int_0^t f(r)dr - \frac{u^2}{2}\int_0^t \sigma^2(r)dr - t\int_{\mathbb{R}}(1 - e^{iuy} + iuy\mathbb{I}_{|y|\leq 1})\nu(dy)\Big), \quad (7.2)$$

where $f(\cdot)$, $\sigma^2(\cdot)$ are functions on $L_1[0, T]$ and ν is a measure on $\mathbb{R}$ satisfying

$$\nu(\{0\}) = 0 \text{ and } \int_{\mathbb{R}}(|y|^2 \wedge 1)\nu(dy) < \infty.$$

In the sequel we shall refer to $(f(\cdot), \sigma^2(\cdot), \nu)$ as the local characteristics of the process $\{X_t\}$ and ν will be called *Lévy measure*. This data characterises uniquely the law of the process $\{X_t\}$. In the case in which $f(\cdot)$ and $\sigma(\cdot)$ are constant functions, a process $\{X_t\}$ satisfying (7.2) is said a *Lévy process* of characteristic triplet (f, σ^2, ν).

Let $D = D([0,\infty), \mathbb{R})$ be the space of mappings ω from $[0,\infty)$ into $\mathbb{R}$ that are right-continuous with left limits. Define the *canonical process* $x : D \to D$ by

$$\forall \omega \in D, \quad x_t(\omega) = \omega_t, \ \ \forall t \geq 0.$$

Let $\mathscr{D}_t$ and $\mathscr{D}$ be the σ-algebras generated by $\{x_s : 0 \leq s \leq t\}$ and $\{x_s : 0 \leq s < \infty\}$, respectively (here, we use the same notations as in Sato (1999)).

Let $\{X_t\}$ be an additive process defined on $(\Omega, \mathscr{A}, \mathbb{P})$ having local characteristics $(f(\cdot), \sigma^2(\cdot), \nu)$. It is well known that it induces a probability measure $P^{(f,\sigma^2,\nu)}$ on (D,$\mathscr{D}$) such that $\big(\{x_t\}, P^{(f,\sigma^2,\nu)}\big)$ is an additive process identical in law with $(\{X_t\}, \mathbb{P})$ (that is the local characteristics of $\{x_t\}$ under $P^{(f,\sigma^2,\nu)}$ is $(f(\cdot), \sigma^2(\cdot), \nu)$). For all $t > 0$ we will denote $P_t^{(f,\sigma^2,\nu)}$ for the restriction of $P^{(f,\sigma^2,\nu)}$ to $\mathscr{D}_t$. In the case where $\int_{|y|\leq 1} |y|\nu(dy) < \infty$, we set $\gamma^\nu := \int_{|y|\leq 1} y\nu(dy)$. Note that, if ν is a finite Lévy measure, then the process $\big(\{x_t\}, P^{(\gamma^\nu,0,\nu)}\big)$ is a compound Poisson process.

Here and in the sequel we will denote by Δx_r the jump of process $\{x_t\}$ at the time r:

$$\Delta x_r = x_r - \lim_{s\uparrow r} x_s.$$

Definition 7.2.2. Consider $\big(\{x_t\}, P^{(f,\sigma^2,\nu)}\big)$ and define the *jump part* of $\{x_t\}$ as

$$x_t^{d,\nu} = \lim_{\varepsilon \to 0} \left(\sum_{r \leq t} \Delta x_r \mathbb{I}_{|\Delta x_r| > \varepsilon} - t \int_{\varepsilon < |y| \leq 1} y\nu(dy) \right) \quad \text{a.s.} \tag{7.3}$$

and its *continuous part* as

$$x_t^{c,\nu} = x_t - x_t^{d,\nu} \quad \text{a.s.} \tag{7.4}$$

We now recall the *Lévy-Itô decomposition*, i.e. the decomposition in continuous and discontinuous parts of an additive process.

Theorem 7.2.3 (See Sato (1999), Theorem 19.3)**.** *Consider* $\big(\{x_t\}, P^{(f,\sigma^2,\nu)}\big)$ *and define* $\{x_t^{d,\nu}\}$ *and* $\{x_t^{c,\nu}\}$ *as in 7.3 and 7.4, respectively. Then the following hold.*

(i) There is $D_1 \in \mathscr{D}$ *with* $P^{(f,\sigma^2,\nu)}(D_1) = 1$ *such that, for any* $\omega \in D_1$, $x_t^{d,\nu}(\omega)$ *is defined for all* $t \in [0,T]$ *and the convergence is uniform in* t *on any bounded interval,* $P^{(f,\sigma^2,\nu)}$*-a.s. The process* $\{x_t^{d,\nu}\}$ *is a Lévy process on* $\mathbb{R}$ *with characteristic triplet* $(0, 0, \nu)$.

(ii) *There is* $D_2 \in \mathscr{D}$ *with* $P^{(f,\sigma^2,\nu)}(D_2) = 1$ *such that, for any* $\omega \in D_2$, $x_t^{c,\nu}(\omega)$ *is continuous in* t. *The process* $\{x_t^{c,\nu}\}$ *is an additive process on* $\mathbb{R}$ *with local characteristics* $(f(\cdot), \sigma^2(\cdot), 0)$.

(iii) *The two processes* $\{x_t^{d,\nu}\}$ *and* $\{x_t^{c,\nu}\}$ *are independent.*

7.2.2 Change of measure for additive processes

For the proof of Theorem 7.1.2 we also need some results on the equivalence of measures for additive processes. By the notation $\ll$ we will mean "is absolutely continuous with respect to".

Case $\sigma^2 = 0$

Theorem 7.2.4 (See Sato (1999), Theorems 33.1–33.2 and Sato (2000) Corollary 3.18, Remark 3.19). *Let* $\big(\{x_t\}, P^{(0,0,\tilde{\nu})}\big)$ *and* $\big(\{x_t\}, P^{(\eta,0,\nu)}\big)$ *be two Lévy processes on* $\mathbb{R}$, *where*

$$\eta = \int_{|y|\le 1} y(\nu - \tilde{\nu})(dy) \tag{7.5}$$

is supposed to be finite. Then $P_t^{(\eta,0,\nu)} \ll P_t^{(0,0,\tilde{\nu})}$ *for all* $t \ge 0$ *if and only if* $\nu \ll \tilde{\nu}$ *and the density* $\frac{d\nu}{d\tilde{\nu}}$ *satisfies*

$$\int \left(\sqrt{\frac{d\nu}{d\tilde{\nu}}(y)} - 1 \right)^2 \tilde{\nu}(dy) < \infty. \tag{7.6}$$

Remark that the finiteness in (7.6) *implies that in* (7.5). *When* $P_t^{(\eta,0,\nu)} \ll P_t^{(0,0,\tilde{\nu})}$, *the density is*

$$\frac{dP_t^{(\eta,0,\nu)}}{dP_t^{(0,0,\tilde{\nu})}}(x) = \exp(U_t(x)),$$

with

$$U_t(x) = \lim_{\varepsilon\to 0} \left(\sum_{r\le t} \ln \frac{d\nu}{d\tilde{\nu}}(\Delta x_r)\mathbb{I}_{|\Delta x_r|>\varepsilon} - \int_{|y|>\varepsilon} t\left(\frac{d\nu}{d\tilde{\nu}}(y) - 1\right)\tilde{\nu}(dy)\right), P^{(0,0,\tilde{\nu})}\text{-a.s.} \tag{7.7}$$

The convergence in (7.7) *is uniform in* t *on any bounded interval,* $P^{(0,0,\tilde{\nu})}$*-a.s. Besides,* $\{U_t(x)\}$ *defined by* (7.7) *is a Lévy process satisfying* $\mathbb{E}_{P^{(0,0,\tilde{\nu})}}[e^{U_t(x)}] = 1$, $\forall t \in [0, T]$.

Case $\sigma^2 > 0$

Lemma 7.2.5. *Let* $\nu_1 \ll \nu_2$ *be Lévy measures such that*

$$\int_{\mathbb{R}} \left(\sqrt{\frac{d\nu_1}{d\nu_2}(y)} - 1 \right)^2 \nu_2(dy) < \infty. \tag{7.8}$$

Define

$$\eta = \int_{|y|\leq 1} y(\nu_1 - \nu_2)(dy), \tag{7.9}$$

which is finite thanks to (7.8)*, and consider real functions* f_1*,* f_2 *and* $\sigma > 0$ *such that*

$$\int_0^T \Big(\frac{f_1(r) - f_2(r) - \eta}{\sigma(r)}\Big)^2 dr < \infty, \quad T \geq 0. \tag{7.10}$$

Then, under $P^{(f_2,\sigma^2,\nu_2)}$*,*

$$M_t(x) = \exp\big(C_t(x) + D_t(x)\big) \tag{7.11}$$

is a $(\mathscr{D}_t)$*-martingale for all* t *in* $[0,T]$*, where*

$$C_t(x) := \int_0^t \frac{f_1(r) - f_2(r) - \eta}{\sigma^2(r)}(dx_r^{c,\nu_2} - f_2(r)dr) - \frac{1}{2}\int_0^t \Big(\frac{f_1(r) - f_2(r) - \eta}{\sigma(r)}\Big)^2 dr,$$

$$D_t(x) := \lim_{\varepsilon\to 0}\left(\sum_{r\leq t} \ln \frac{d\nu_1}{d\nu_2}(\Delta x_r)\mathbb{I}_{|\Delta x_r|>\varepsilon} - t\int_{|y|>\varepsilon}(\nu_1 - \nu_2)(dy)\right). \tag{7.12}$$

The convergence in (7.12) *is uniform in* t *on any bounded interval,* $P^{(f_2,\sigma^2,\nu_2)}$*-a.s.*

Proof. The existence of the limit in (7.12) is guaranteed by (7.8) (see Theorem 7.2.4). Since $\int_0^t \frac{1}{\sigma(r)}(dx_r^{c,\nu_2} - f_2(r)dr)$ is a standard Brownian motion under $P^{(f_2,\sigma^2,0)}$, we have that $\int_s^t \frac{f_1(r)-f_2(r)-\eta}{\sigma^2(r)}(dx_r^{c,\nu_2} - f_2(r)dr)$ has normal law $\mathscr{N}\Big(0, \int_s^t \big(\frac{f_1(r)-f_2(r)-\eta}{\sigma(r)}\big)^2 dr\Big)$, hence $\mathbb{E}_{P^{(f_2,\sigma^2,0)}}[\exp((C_t - C_s)(x))] = 1$. Theorem 7.2.3 entails that $\{x_t^{c,\nu_2}\}$ and $\{x_t^{d,\nu_2}\}$ are independent under $P^{(f_2,\sigma^2,\nu_2)}$. Moreover, the law of $\{C_t(x)\}$ (resp. $\{D_t(x)\}$) is the same under $P^{(f_2,\sigma^2,\nu_2)}$ or $P^{(f_2,\sigma^2,0)}$ (resp. $P^{(f_2,0,\nu_2)}$ or $P^{(0,0,\nu_2)}$). Further, using Theorem 7.2.4, we know that $\{D_t(x)\}$ is a Lévy process such that $\mathbb{E}_{P^{(0,0,\nu_2)}}[\exp(D_{t-s}(x))] = 1$ for all $s < t$. These facts together with the independence of the increments of $(\{x_t\}, P^{(f_2,\sigma^2,\nu_2)})$ and the stationarity of $\{D_t(x)\}$ imply:

$$
\begin{aligned}
\mathbb{E}_{P(f_2,\sigma^2,\nu_2)}[M_t(x)|\mathscr{D}_s] &= \mathbb{E}_{P(f_2,\sigma^2,\nu_2)}\Big[M_s(x)\exp\Big((C_t-C_s)(x)+(D_t-D_s)(x)\Big)\Big|\mathscr{D}_s\Big] \\
&= M_s(x)\mathbb{E}_{P(f_2,\sigma^2,\nu_2)}[\exp((C_t-C_s)(x)+(D_t-D_s)(x))] \\
&= M_s(x)\mathbb{E}_{P(f_2,\sigma^2,0)}[\exp((C_t-C_s)(x))]\mathbb{E}_{P(0,0,\nu_2)}[\exp((D_t-D_s)(x))] \\
&= M_s(x)\mathbb{E}_{P(0,0,\nu_2)}[\exp(D_{t-s}(x))] \\
&= M_s(x).
\end{aligned}
$$

□

Lemma 7.2.6. *Suppose that the hypothesis* (7.8) *and* (7.10) *of Lemma 7.2.5 are satisfied. Then, using the same notations as above,* $P_t^{(f_1,\sigma^2,\nu_1)} \ll P_t^{(f_2,\sigma^2,\nu_2)}$ *for all* t *and the density is given by:*

$$
\frac{dP_t^{(f_1,\sigma^2,\nu_1)}}{dP_t^{(f_2,\sigma^2,\nu_2)}}(x) = M_t(x). \tag{7.13}
$$

Proof. For $s<t$, we prove that

$$
\mathbb{E}_{P(f_2,\sigma^2,\nu_2)}\big[\exp(iu(x_t-x_s))\frac{M_t}{M_s}(x)|\mathscr{D}_s\big] = \mathbb{E}_{P(f_1,\sigma^2,\nu_1)}[\exp(iu(x_t-x_s)].
$$

To that aim remark that, thanks again to Theorem 7.2.3:

$$
\begin{aligned}
\mathbb{E}_{P(f_2,\sigma^2,\nu_2)}\Big[e^{iu(x_t-x_s)}\frac{M_t(x)}{M_s(x)}\Big|\mathscr{D}_s\Big] &= \mathbb{E}_{P(f_2,\sigma^2,\nu_2)}\Big[e^{iu(x_t^{c,\nu_2}-x_s^{c,\nu_2}+x_t^{d,\nu_2}-x_s^{d,\nu_2})}\frac{M_t(x)}{M_s(x)}\Big|\mathscr{D}_s\Big] \\
&= \mathbb{E}_{P(f_2,\sigma^2,0)}\Big[e^{iu(x_t-x_s)}e^{(C_t-C_s)(x)}\Big]\mathbb{E}_{P(0,0,\nu_2)}\Big[e^{iu(x_t-x_s)}e^{(D_t-D_s)(x)}\Big].
\end{aligned} \tag{7.14}
$$

Let us now compute the first factor of (7.14):

$$
\begin{aligned}
\mathbb{E}_{P(f_2,\sigma^2,0)}\Big[e^{iu(x_t-x_s)}e^{(C_t-C_s)(x)}\Big] &= \mathbb{E}_{P(f_1-\cdot\eta,\sigma^2,0)}\Big[e^{iu(x_t-x_s)}\Big] \\
&= \exp\Big(iu\int_s^t (f_1(r)-\eta)dr - \frac{u^2}{2}\int_s^t \sigma^2(r)dr\Big).
\end{aligned}
$$

In the first equality we used the Girsanov theorem, thanks to the fact that $\int_0^t \frac{1}{\sigma(r)}(dx_r - f_2(r)dr)$ is a Brownian motion under $P^{(f_2,\sigma^2,0)}$, while the second one follows from (7.2). We compute the second factor of (7.14) by means of Theorem 7.2.4 and another application of (7.2):

$$
\begin{aligned}
\mathbb{E}_{P(0,0,\nu_2)}\Big[e^{iu(x_t-x_s)}e^{(D_t-D_s)(x)}\Big] &= \mathbb{E}_{P(0,0,\nu_2)}\Big[e^{iux_{t-s}}e^{D_{t-s}(x)}\Big] \\
&= \mathbb{E}_{P(\eta,0,\nu_1)}\Big[e^{iux_{t-s}}\Big] \\
&= \exp\Big((t-s)\Big[iu\eta - \int_{\mathbb{R}}(1-e^{iuy}+iuy\mathbb{I}_{|y|\le 1})\nu_1(dy)\Big]\Big).
\end{aligned}
$$

Consequently:

$$\mathbb{E}_{P(f_2,\sigma^2,\nu_2)}\left[e^{iu(x_t-x_s)}\frac{M_t(x)}{M_s(x)}\Big|\mathscr{D}_s\right]=\mathbb{E}_{P(f_1,\sigma^2,\nu_1)}[e^{iu(x_t-x_s)}]\quad \forall 0\le s\le t. \tag{7.15}$$

Fix t and define a probability measure P_t on $\mathscr{D}_t$ by $P_t(B)=\mathbb{E}_{P(f_2,\sigma^2,\nu_2)}[M_t\mathbb{I}_B]$ for $B\in\mathscr{D}_t$. As a consequence of Lemma 7.2.5 and the Bayes rule, the two processes given by $\big(\{x_s: 0\le s\le t\},P_t^{(f_1,\sigma^2,\nu_1)}\big)$ and $\big(\{x_s:0\le s\le t\},P_t\big)$ are identical. Indeed, by (7.15), both have independent increments and the prescribed characteristic function. Consequently, (7.13) holds. □

7.3 Proof of Theorem 7.2.2

For the proof we will need the following three calculus lemmas.

Lemma 7.3.1. *Let X be a random variable with normal law $\mathcal{N}(m,\sigma^2)$. Then*

$$\mathbb{E}\Big|1-e^X\Big|=2\Big[\phi\Big(-\frac{m}{\sigma}\Big)-\phi\Big(-\frac{m}{\sigma}-\sigma\Big)\Big],$$

where $\phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-\frac{y^2}{2}}dy$.

Proof. By definition we have

$$\begin{aligned}\mathbb{E}\Big|1-e^X\Big|&=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}|1-e^x|e^{-\frac{(x-m)^2}{2\sigma^2}}dx\\&=\frac{1}{\sqrt{2\pi}\sigma}\Big(\int_{-\infty}^{0}(1-e^x)e^{-\frac{(x-m)^2}{2\sigma^2}}dx+\int_0^{\infty}(e^x-1)e^{-\frac{(x-m)^2}{2\sigma^2}}dx\Big).\end{aligned}$$

To conclude, just split the sums inside the integrals and use the change of variables $\big(y=\frac{x-m}{\sigma}-\sigma\big)$, resp. $\big(y=\frac{x-m}{\sigma}\big)$. □

Lemma 7.3.2. *For all x,y in $\mathbb{R}$ we have:*

$$|1-e^{x+y}|\le\frac{1+e^x}{2}|1-e^y|+\frac{1+e^y}{2}|1-e^x|. \tag{7.16}$$

Proof. By symmetry we restrict to $x\ge 0$.

- $x,y\ge 0$: In this case we have that $|1-e^{x+y}|$ is exactly equal to $\frac{1+e^x}{2}|1-e^y|+\frac{1+e^y}{2}|1-e^x|$.

- $x \geq 0, y \leq 0, x+y \geq 0$: Then the member on the right hand side of (7.16) is equal to $e^x - e^y \geq e^x - 1 \geq e^{x+y} - 1$.
- $x \geq 0, y \leq 0, x+y \leq 0$: In this case the member on the right of (7.16) is equal to $e^x - e^y \geq 1 - e^y \geq 1 - e^{x+y}$.

$\square$

Lemma 7.3.3. *With the same notations as in Theorem 7.1.2 and Lemma 7.2.5, we have:*

$$\mathbb{E}_{P_T^{(0,0,\nu_2)}}\Big[\big|1-\exp(D_T(x))\big|\Big] = \mathbb{E}_{P_T^{(\gamma^{\nu_2},0,\nu_2)}}\Big[\big|1-\exp(D_T(x))\big|\Big] \leq 2\sinh\left(T\int_{\mathbb{R}} L_1(\nu_1,\nu_2)\right). \tag{7.17}$$

Proof. Thanks to Theorem 7.2.3 it is clear that $\mathbb{E}_{P_T^{(0,0,\nu_2)}}\big[|1-\exp(D_T(x))|\big] = \mathbb{E}_{P_T^{(\gamma^{\nu_2},0,\nu_2)}}\big[|1-\exp(D_T(x))|\big]$. In order to simplify the notations, let us introduce the quantities $h^+ = \left(\frac{d\nu_1}{d\nu_2}\right)^{\mathbb{I}_{\frac{d\nu_1}{d\nu_2}\geq 1}}$ and $h^- = \left(\frac{d\nu_1}{d\nu_2}\right)^{\mathbb{I}_{\frac{d\nu_1}{d\nu_2}< 1}}$ (i.e. h^+, resp. h^-, is identically 1 where $\frac{d\nu_1}{d\nu_2} < 1$, resp. where $\frac{d\nu_1}{d\nu_2} \geq 1$). Let us define

$$A^+(x) := \lim_{\varepsilon\to 0}\left(\sum_{r\leq T}\ln h^+(\Delta x_r)\mathbb{I}_{|\Delta(x_r)|>\varepsilon} - T\int_{|y|>\varepsilon}(h^-(y)-1)\nu_2(dy)\right),$$
$$A^-(x) := \lim_{\varepsilon\to 0}\left(\sum_{r\leq T}\ln h^-(\Delta x_r)\mathbb{I}_{|\Delta(x_r)|>\varepsilon} - T\int_{|y|>\varepsilon}(h^+(y)-1)\nu_2(dy)\right).$$

Remark that $A^+(x)$ and $A^-(x)$ have the same law under $P_T^{(\gamma^{\nu_2},0,\nu_2)}$ and $P_T^{(0,0,\nu_2)}$ (see Theorem 7.2.3), and that they are constructed in such a way that

$$D_T(x) = A^+(x) + A^-(x).$$

Using Lemma 7.3.2 and the fact that $A^+(x) \geq 0$ and $A^-(x) \leq 0$ we get:

$$\begin{aligned}\mathbb{E}_{P_T^{(\gamma^{\nu_2},0,\nu_2)}}\big[|1-D_T(x)|\big] &= \mathbb{E}_{P_T^{(\gamma^{\nu_2},0,\nu_2)}}\big|1-\exp(A^+(x)+A^-(x))\big| \\ &\leq \mathbb{E}_{P_T^{(\gamma^{\nu_2},0,\nu_2)}}\left[\frac{1+e^{A^+(x)}}{2}\Big|1-e^{A^-(x)}\Big| + \frac{1+e^{A^-(x)}}{2}\Big|1-e^{A^+(x)}\Big|\right] \\ &= \mathbb{E}_{P_T^{(\gamma^{\nu_2},0,\nu_2)}}\Big[e^{A^+(x)} - e^{A^-(x)}\Big].\end{aligned}$$

In order to compute the last quantity let us consider two more Lévy measures:

$$\nu^+ := \nu_2\mathbb{I}_{\frac{d\nu_1}{d\nu_2}<1} + \nu_1\mathbb{I}_{\frac{d\nu_1}{d\nu_2}\geq 1}, \quad \nu^- := \nu_2\mathbb{I}_{\frac{d\nu_1}{d\nu_2}\geq 1} + \nu_1\mathbb{I}_{\frac{d\nu_1}{d\nu_2}<1}.$$

Notice that ν^+, ν^- are absolutely continuous with respect to ν_2, with densities given by h^+, h^-, respectively. Applying Theorem 7.2.4 to the pairs of measure (ν^+, ν_2) and (ν^-, ν_2) we get

$$\mathbb{E}_{P_T^{(0,0,\nu_2)}}\left[e^{A^+(x)}\right] = \exp\left(T\int_{\mathbb{R}}(h^+(y) - h^-(y))\nu_2(dy)\right),$$
$$\mathbb{E}_{P_T^{(0,0,\nu_2)}}\left[e^{A^-(x)}\right] = \exp\left(T\int_{\mathbb{R}}(h^-(y) - h^+(y))\nu_2(dy)\right).$$

Then, recalling that both $A^+(x)$ and $A^-(x)$ have the same law under $P_T^{(\gamma^{\nu_2},0,\nu_2)}$ and $P_T^{(0,0,\nu_2)}$ we obtain:

$$\mathbb{E}_{P_T^{(\gamma^{\nu_2},0,\nu_2)}}\left[e^{A^+(x)} - e^{A^-(x)}\right] = 2\sinh\left(T\int_{\mathbb{R}}\left|1 - \frac{d\nu_1}{d\nu_2}(y)\right|\nu_2(dy)\right).$$

$\square$

Proof of Theorem 7.1.2. Case $\sigma^2 > 0$*:* With the same notations as in Lemma 7.2.5 and by means of Lemma 7.2.6 one can write

$$L_1\big(P_T^{(f_1,\sigma^2,\nu_1)}, P_T^{(f_2,\sigma^2,\nu_2)}\big) = \mathbb{E}_{P_T^{(f_2,\sigma^2,\nu_2)}}\big|1 - \exp(C_T(x) + D_T(x))\big|.$$

Now, using Lemma 7.3.2 and the independence between $C_T(x)$ and $D_T(x)$ (Theorem 7.2.3), we obtain

$$\begin{aligned} L_1\big(P_T^{(f_2,\sigma^2,\nu_2)}, P_T^{(f_1,\sigma^2,\nu_1)}\big) \le & \mathbb{E}_{P_T^{(f_2,\sigma^2,\nu_2)}}\left(\frac{1+e^{C_T(x)}}{2}\right)\mathbb{E}_{P_T^{(f_2,\sigma^2,\nu_2)}}|1 - e^{D_T(x)}| \\ & + \mathbb{E}_{P_T^{(f_2,\sigma^2,\nu_2)}}\left(\frac{1+e^{D_T(x)}}{2}\right)\mathbb{E}_{P_T^{(f_2,\sigma^2,\nu_2)}}|1 - e^{C_T(x)}|. \end{aligned}$$

We conclude the proof using Lemmas 7.3.3 and 7.3.1 together with the fact that

$$\mathbb{E}_{P_T^{(f_2,\sigma^2,\nu_2)}}e^{C_T(x)} = 1 = \mathbb{E}_{P_T^{(f_2,\sigma^2,\nu_2)}}e^{D_T(x)}.$$

Case $\sigma^2 = 0$*:* If $f_1 - f_2 \equiv \gamma^{\nu_1} - \gamma^{\nu_2}$, notice that, as the drift component of $\big(\{x_t\}, P_T^{(f_1,0,\nu_1)}\big)$ and $\big(\{x_t\}, P_T^{(f_2,0,\nu_2)}\big)$ is deterministic, we have

$$\frac{dP_T^{(f_1,0,\nu_1)}}{dP_T^{(f_2,0,\nu_2)}}(x) = \frac{dP_T^{(f_1-f_2,0,\nu_1)}}{dP_T^{(0,0,\nu_2)}}(x) = D_T(x)$$

with $D_T(x)$ as in (7.12). Theorem 7.2.4 allows us to write the L_1-distance between $P_T^{(f_1,0,\nu_1)}$ and $P_T^{(f_2,0,\nu_2)}$ as $\mathbb{E}_{P_T^{(f_2,0,\nu_2)}}\big|1 - D_T(x)\big|$. We then obtain the bound $2\sinh(TL_1(\nu_1,\nu_2))$ by means of Lemma 7.3.3. $\square$

Chapitre 8

Conclusions et perspectives

8.1 Conclusion générale

Cette thèse se situe à la frontière entre la statistique mathématique et l'inférence pour des processus stochastiques. Dans notre travail, la théorie de Le Cam sur la comparaison d'expériences statistiques a été appliquée à plusieurs modèles liés à l'observation discrète à haute fréquence d'une trajectoire d'un processus stochastique. Nous avons étudié le cas d'une diffusion, d'un processus additif et plus particulièrement d'un processus de Lévy. L'objectif était d'utiliser la distance de Le Cam pour avoir une compréhension plus profonde des différents problèmes statistiques liés à ces processus.

Tout d'abord, nous nous sommes concentrés sur des modèles dirigés par des processus à accroissements indépendants et à trajectoires càdlàg. À notre connaissance, il s'agit des premiers résultats d'équivalence asymptotique pour les processus à sauts. Notre premier résultat dans ce contexte est l'équivalence asymptotique globale entre l'expérience engendrée par l'observation discrète ou continue d'un processus additif et d'un bruit blanc gaussien. Ici, le paramètre d'intérêt est la fonction de dérive. Ensuite, nous avons considéré le problème de l'estimation de la densité de Lévy lorsque l'on ne dispose que d'observations discrètes d'un processus de Lévy à sauts purs. Ce problème s'avère asymptotiquement équivalent à l'estimation de la dérive d'un certain modèle de bruit blanc gaussien. Toutes les équivalences asymptotiques ont été établies en construisant des noyaux de Markov explicites qui peuvent être utilisés pour reproduire une expérience à partir de l'autre. Ces travaux sur les processus à accroissements indépendants ont donné lieu à deux articles Mariucci 2015b,d.

Les outils que nous avons développés dans cette première partie nous ont permis d'affaiblir de manière considérable les hypothèses utilisées pour prouver le résultat d'équivalence asymptotique entre l'estimation de la densité f issue de n variables aléatoires i.i.d. et celle de la dérive $\sqrt{f}$ d'un bruit blanc gaussien de variance $1/4n$. Ce résultat a fait l'objet d'un article (voir Mariucci 2015a).

Après avoir traité le cas des processus à accroissements indépendants nous avons considéré des modèles liés à des processus de diffusion, en utilisant pour ce cas des techniques très différentes. Plus précisément, nous avons prouvé l'équivalence asymptotique globale entre les modèles de diffusion unidimensionnels de dérive inconnue et en présence de petite variance et les modèles autorégressifs non paramétriques. Nous avons traité à la fois les cas des observations discrètes et continues. Ces résultats sont publiés dans Mariucci 2015c.

Travailler avec la distance de Le Cam impose de bien contrôler la distance en variation totale entre les lois associées aux modèles d'intérêt. Ceci nous a amené à nous intéresser à la distance L_1 entre les lois des processus additifs. Nous en avons proposé une majoration qui ne dépend que de la connaissance de la caractéristique locale des processus. Ce travail est détaillé dans une publication Étoré, Mariucci 2014.

8.2 Extensions possibles des travaux de thèse

Différentes perspectives de recherche s'ouvrent à l'issue de cette thèse. Nous en discutons ici quatre d'entre elles que nous nous proposons de développer dans de futurs travaux.

8.2.1 Estimation non paramétrique de la densité de Lévy

La dynamique des sauts d'un processus de Lévy X est entièrement dictée par sa densité de Lévy. Comprendre le comportement des sauts nécessite donc d'estimer la densité de Lévy. Plusieurs travaux récents ont traité ce problème, voir Belomestny et al. 2015 pour une présentation détaillée. Cependant, ce problème est difficile mathématiquement, car, dans la pratique, nous n'avons accès qu'à des observations discrètes $X_{t_0}, \dots, X_{t_n}$ pour lesquelles, en général, nous ne sommes pas capables de caractériser la loi de manière satisfaisante.

Une possible direction de recherche consisterait alors à utiliser le résultat d'équivalence prouvé dans Mariucci 2015d pour construire un estimateur de la densité de Lévy. Jusqu'à présent, les résultats dans Mariucci 2015d sont non-constructifs au sens où ils ne fournissent pas une procédure explicite pour construire un estimateur de la densité de Lévy f dans l'expérience liée aux observations discrètes de X à partir de la connaissance d'un estimateur de la dérive $\sqrt{f}$ du modèle, plus simple, du bruit blanc gaussien. Cependant, les équivalences présentées dans Mariucci 2015d ont été obtenues en construisant des noyaux markoviens explicites : un travail de ré-écriture de la preuve nous garantirait la construction d'un estimateur minimax de la densité de Lévy. Remarquons ici qu'une approche similaire pour la construction de procédures statistiques en utilisant la théorie de Le Cam a déjà été exploitée avec succès, par exemple, dans Brown, Low 1996 (voir le Corollaire 4.1) pour construire un estimateur de la fonction de régression et dans Dalalyan, Reiß 2006 (voir la Section 4.1) pour un estimateur de la dérive d'un processus de diffusion.

Concrètement, une perspective directe découlant du travail présenté dans le Chapitre 4 est la suivante. Nous considérons un processus à sauts X, sans partie gaussienne et avec une mesure de Lévy ν. Notons par f sa densité de Lévy par rapport à une certaine mesure de référence ν_0 (supposée connue) et supposons que f appartienne à une certaine classe fonctionnelle $\mathscr{F}$. Le but est alors d'estimer f à partir des observations de la forme $X_{i\Delta_n}$, $i = 0, 1, \dots, T_n/\Delta_n$, avec $T_n \to \infty$ et $\Delta_n \to 0$ lorsque $n \to \infty$. Remarquons aussi que dans le résultat d'équivalence Mariucci 2015d nous avons traité le cas de mesures de Lévy éventuellement infinie et pas nécessairement à variation finie. De plus, le résultat était valable dans une fenêtre d'observation de f qui pouvait être infinie et contenir le point critique 0. Nous pensons donc qu'il est possible de construire un estimateur minimax de f sous les mêmes conditions.

8.2.2 Équivalence asymptotique pour des modèles de Lévy dépendant aléatoirement du temps

Les processus de Lévy reçoivent beaucoup d'attention depuis de nombreuses années en raison de leur simplicité mathématique d'une part et de leur capacité à reproduire de nombreuses propriétés spécifiques des données économiques d'autre part. Récemment, des modèles basés sur les processus de Lévy ont été proposés pour mieux tenir compte des caractéristiques du prix des options. Les modèles principaux sont les processus de Lévy dépendant aléatoirement du temps (time-changed Lévy processes). Un processus de Lévy dépendant aléatoirement du temps, Y, peut être représenté sous la forme

$$Y_s = X_{\tau_s},$$

où X est un processus de Lévy et τ est un processus stochastique positive et croissant (voir, par exemple, Carr, Wu 2004 pour l'application des processus en finance).

Comme déjà souligné, les résultats d'équivalence au sens de Le Cam fournissent une compréhension plus profonde des problèmes statistiques et permettent de mettre en lumière leurs principales caractéristiques. Cependant, il y a peu de résultats d'équivalence concernant les processus à sauts et une ligne de recherche intéressante pourrait être celle d'étendre le résultat d'équivalence dans Mariucci 2015d à des situations plus générales, comprenant des modèles de Lévy dépendant aléatoirement du temps.

8.2.3 Équivalence asymptotique pour des modèles de diffusion multidimensionnels avec un coefficient de diffusion inconnu

Dans l'article Mariucci 2015c nous avons démontré, à travers des techniques de changement de temps pour des processus en temps continu, l'équivalence asymptotique entre un processus de diffusion observé en haute fréquence et le schéma d'Euler correspondant. Dans ce contexte, le paramètre inconnu est la fonction de dérive alors que le coefficient de diffusion $\varepsilon\sigma(\cdot)$ est supposé connu (mais pas constant).

Au moins deux directions d'extension se proposent : tout d'abord, il serait très intéressant de comprendre si, en supprimant l'hypothèse $\varepsilon\sigma(\cdot)$ connu, le résultat d'équivalence reste valable. Dans la pratique, en effet, généralement on ne connaît pas $\varepsilon\sigma(\cdot)$ qui est habituellement considéré comme un paramètre de nuisance secondaire. Deuxièmement, des résultats d'équivalence pour des modèles de diffusion en dimension supérieure ayant des coefficients de diffusion non constants ne sont pas encore disponibles. En fait, des résultats d'équivalence asymptotique dans le cas multidimensionnel sont très rares en général, même si les modèles multidimensionnels sont couramment utilisés dans les applications (voir par exemple Aït-Sahalia 2008 pour une discussion sur les applications des processus de diffusion multidimensionnels en économétrie). Une perspective de recherche serait alors de développer de nouvelles méthodes et outils afin de gérer un résultat d'équivalence concernant l'inférence statistique de la dérive dans une expérience de diffusion multidimensionnelle.

Plus concrètement, un résultat clé pour prouver l'équivalence asymptotique entre un processus de diffusion observé d'une manière discrète et son schéma d'Euler sans assumer la connaissance du coefficient de diffusion pourrait être l'article de Carter 2007. Dans ce travail, l'auteur a généralisé l'équivalence bien connue entre un modèle de régression non paramétrique et un modèle de bruit blanc gaussien au cas où la variance est inconnue. En ce qui concerne une possible extension en dimension supérieure, on peut constater qu'il y a encore équivalence asymptotique dans certains cas particuliers et en dimension 2 (Brownian oscillator), mais, afin d'obtenir une généralisation d-dimensionnelle, une analyse plus profonde est nécessaire. Par exemple, un point de vue intéressant sur le sujet pourrait être celui de Strauch 2015.

8.2.4 Évaluation de la distance entre les lois des accroissements des processus de Lévy

Donner des majorations pour la distance de la variation totale (mais aussi pour la distance de Hellinger et la divergence de Kullback-Leibler) est un problème classique, qui, dans les dernières décennies, a été réinterprété en termes plus modernes via la méthode de Stein.

En particulier, lorsqu'il s'agit de processus de Lévy observés de manière discrète, le problème de la comparaison de leurs incréments apparaît souvent. Plus précisément, soient X^i, $i=1,2$, deux processus de Lévy de triplet (b_i,σ_i^2,ν_i). D'un point de vue technique, il serait utile de contrôler les quantités suivantes :

$$\|\mathcal{L}(X_t^1)-\mathcal{L}(X_t^2)\|_{TV},\quad H(\mathcal{L}(X_t^1)-\mathcal{L}(X_t^2)),\quad D(\mathcal{L}(X_t^1)|\mathcal{L}(X_t^2)),$$

où $\mathcal{L}$ signifie “la loi de”, $\|\cdot\|_{TV}$, $H(\cdot,\cdot)$ et $D(\cdot|\cdot)$ dénotent “distance en variation totale”, “distance d'Hellinger” et “divergence de Kullback-Leibler”, respectivement.

Certains résultats concernant la distance en variation totale ou la distance d'Hellinger entre les trajectoires des processus de Lévy existent déjà (voir par exemple Jacod, Shiryaev 1987, Liese 1987 et Étoré, Mariucci 2014) et ils peuvent être utilisés pour dériver des bornes de la distance entre les incréments des processus. Par exemple, un résultat classique présenté dans Jacod, Shiryaev 1987 est que si X^i, $i=1,2$ sont des processus de Lévy à sauts purs, alors la distance d'Hellinger entre la loi des processus X^1 et X^2 observés jusqu'à l'instant T est majorée par :

$$H^2\Big(\mathcal{L}\big((X_t^1)_{t\in[0,T]}\big),\mathcal{L}\big((X_t^2)_{t\in[0,T]}\big)\Big)\leq\frac{T}{2}H^2(\nu_1,\nu_2).$$

Sachant que les noyaux markoviens réduisent la distance en variation totale, on peut utiliser le résultat ci-dessus pour déduire que :

$$\frac{H^2(\mathcal{L}(X_t^1),\mathcal{L}(X_t^2))}{2}\leq\|\mathcal{L}(X_t^1)-\mathcal{L}(X_t^2)\|_{TV}\leq\sqrt{\frac{t}{2}}H(\nu_1,\nu_2).$$

Cependant, on devrait être en mesure d'obtenir des limites plus fines, puisque l'ordre de grandeur de la dernière équation semble être sous optimal. Obtenir de tels résultats permettrait de développer des outils importants pour majorer la distance de Le Cam entre des modèles associés à l'observation discrète d'un processus de Lévy.

Bibliographie

Aït-Sahalia, Y. (2008). 'Closed-form likelihood expansions for multivariate diffusions'. In: *The Annals of Statistics* 36.2, pp. 906–937.

Bec, M., C. Lacour (2012). 'Adaptive kernel estimation of the Lévy density'. In: *Hal preprint* 583221.

Belomestny, D., F. Comte, V. Genon-Catalot, H. Masuda, M. Reiß (2015). *Lévy Matters IV*.

Bickel, P. J., C. A. Klaassen, P. J. Bickel, Y Ritov, J Klaassen, J. A. Wellner, Y. Ritov (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press Baltimore.

Billingsley, P. (1999). *Convergence of probability measures*. Second. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. New York: John Wiley & Sons Inc., pp. x+277.

Blackwell, D. (1947). 'Conditional expectation and unbiased sequential estimation'. In: *Ann. Math. Statistics* 18, pp. 105–110.

— (1951). 'Comparison of experiments'. In: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*. Berkeley and Los Angeles: University of California Press, pp. 93–102.

— (1953). 'Equivalent comparisons of experiments'. In: *Ann. Math. Statistics* 24, pp. 265–272.

Bohnenblust, A., L. Shapley, S. Sherman (1949). 'Reconnaissance in game theory'. In: *Rand Research Memorandum 1949/208*.

Brockett, P. L., H. G. Tucker (1977). 'A conditional dichotomy theorem for stochastic processes with independent increments'. In: *J. Multivariate Anal.* 7.1, pp. 13–27.

Brown, L. D., M. G. Low (1996). 'Asymptotic equivalence of nonparametric regression and white noise'. In: *Ann. Statist.* 24.6, pp. 2384–2398.

Brown, L. D., Y. Wang, L. H. Zhao (2003). 'On the statistical equivalence at suitable frequencies of GARCH and stochastic volatility models with the corresponding diffusion model'. In: *Statist. Sinica* 13.4. Statistical applications in financial econometrics, pp. 993–1013.

Brown, L. D., C.-H. Zhang (1998). 'Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2'. In: *Ann. Statist.* 26.1, pp. 279–287.

Brown, L. D., T. T. Cai, M. G. Low, C.-H. Zhang (2002a). 'Asymptotic equivalence theory for nonparametric regression with random design'. In: *Ann. Statist.* 30.3. Dedicated to the memory of Lucien Le Cam, pp. 688–707.

— (2002b). 'Asymptotic equivalence theory for nonparametric regression with random design'. In: *Ann. Statist.* 30.3. Dedicated to the memory of Lucien Le Cam, pp. 688–707.

Brown, L. D., A. V. Carter, M. G. Low, C.-H. Zhang (2004a). 'Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift'. In: *Ann. Statist.* 32.5, pp. 2074–2097.

— (2004b). 'Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift'. In: *Ann. Statist.* 32.5, pp. 2074–2097.

Buchmann, B., G. Müller (2012). 'Limit experiments of GARCH'. In: *Bernoulli* 18.1, pp. 64–99.

Buscemi, F. (2012). 'Comparison of quantum statistical models: equivalent conditions for sufficiency'. In: *Comm. Math. Phys.* 310.3, pp. 625–647.

Cai, T. T., H. H. Zhou (2009). 'Asymptotic equivalence and adaptive estimation for robust nonparametric regression'. In: *Ann. Statist.* 37.6A, pp. 3204–3235.

Carr, P., L. Wu (2004). 'Time-changed Lévy processes and option pricing'. In: *Journal of Financial economics* 71.1, pp. 113–141.

Carr, P., H. Geman, D. B. Madan, M. Yor (2002). 'The fine structure of asset returns: An empirical investigation'. In: *The Journal of Business* 75.2, pp. 305–333.

Carter, A., D. Pollard (2004). 'Tusnády's inequality revisited'. In: *The Annals of Statistics* 32.6, pp. 2731–2741.

Carter, A. V. (2002). 'Deficiency distance between multinomial and multivariate normal experiments'. In: *Ann. Statist.* 30.3. Dedicated to the memory of Lucien Le Cam, pp. 708–730.

— (2006a). 'A continuous Gaussian approximation to a nonparametric regression in two dimensions'. In: *Bernoulli* 12.1, pp. 143–156.

— (2006b). 'A continuous Gaussian approximation to a nonparametric regression in two dimensions'. In: *Bernoulli* 12.1, pp. 143–156.

— (2007). 'Asymptotic approximation of nonparametric regression experiments with unknown variances'. In: *Ann. Statist.* 35.4, pp. 1644–1673.

— (2009). 'Asymptotically sufficient statistics in nonparametric regression experiments with correlated noise'. In: *J. Probab. Stat.* Art. ID 275308, 19.

Comte, F., V. Genon-Catalot (2010). 'Nonparametric adaptive estimation for pure jump Lévy processes'. In: *Ann. Inst. H. Poincaré Probab. Statist* 46.3, pp. 595–617.

— (2011). 'Estimation for Lévy processes from high frequency data within a long time interval'. In: *The Annals of Statistics* 39.2, pp. 803–837.

Comte, F., V. Genon-Catalot, Y. Rozenholc (2007). 'Penalized nonparametric mean square estimation of the coefficients of diffusion processes'. In: *Bernoulli*, pp. 514–543.

Cont, R., P. Tankov (2004). *Financial modelling with jump processes.* Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, xvi+535.

Cover, T. M., J. A. Thomas (2012). *Elements of information theory.* John Wiley & Sons.

Dalalyan, A., M. Reiß (2006). 'Asymptotic statistical equivalence for scalar ergodic diffusions'. In: *Probab. Theory Related Fields* 134.2, pp. 248–282.

— (2007a). 'Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case'. In: *Probab. Theory Related Fields* 137.1-2, pp. 25–47.

— (2007b). 'Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case'. In: *Probab. Theory Related Fields* 137.1-2, pp. 25–47.

Delattre, S., M. Hoffmann (2002). 'Asymptotic equivalence for a null recurrent diffusion'. In: *Bernoulli* 8.2, pp. 139–174.

Diaconis, P., S. L. Zabell (1982). 'Updating subjective probability'. In: *Journal of the American Statistical Association* 77.380, pp. 822–830.

Diebolt, J., L. Gardes, S. Girard, A. Guillou (2008). 'Bias-reduced estimators of the Weibull tail-coefficient'. In: *TEST* 17.2, pp. 311–331.

Duval, C. (2013). 'Density estimation for compound Poisson processes from discrete data'. In: *Stochastic Process. Appl.* 123.11, pp. 3963–3986.

Duval, C., M. Hoffmann (2011). 'Statistical inference across time scales'. In: *Electron. J. Stat.* 5, pp. 2004–2030.

Efromovich, S., A. Samarov (1996). 'Asymptotic equivalence of nonparametric regression and white noise model has its limits'. In: *Statist. Probab. Lett.* 28.2, pp. 143–145.

Etoré, P., S. Louhichi, E. Mariucci (2013). *Asymptotic equivalence of jumps Lévy processes and their discrete counterpart.*

Étoré, P., E. Mariucci (2014). 'L1-distance for additive processes with time-homogeneous Lévy measures'. In: *Electronic Communications in Probability* 19, pp. 1–10.

Figueroa-López, J. E. (2009). 'Nonparametric estimation for Lévy models based on discrete-sampling'. In: *Lecture Notes-Monograph Series*, pp. 117–146.

Figueroa-López, J. E., C. Houdré (2006). 'Risk bounds for the non-parametric estimation of Lévy processes'. In: *Lecture Notes-Monograph Series*, pp. 96–116.

Freidlin, M. I., A. D. Wentzell (2012). *Random perturbations of dynamical systems*. Third. Vol. 260. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Translated from the 1979 Russian original by Joseph Szücs. Springer, Heidelberg, pp. xxviii+458.

Gairing, J., M. Högele, T. Kosenkova, A. Kulik (2013). 'Coupling distances between Lévy measures and applications to noise sensitivity of SDE'. In: *Stochastics and Dynamics*.

Gardes, L., S. Girard (2006). 'Comparison of Weibull tail-coefficient estimators'. In: *REV-STAT* 4.2, pp. 163–188.

Genon-Catalot, V (1990). 'Maximum contrast estimation for diffusion processes from discrete observations'. In: *Statistics* 21.1, pp. 99–116.

Genon-Catalot, V., C. Laredo (2014). 'Asymptotic equivalence of nonparametric diffusion and Euler scheme experiments'. In: *The Annals of Statistics* 42.3, pp. 1145–1165.

Genon-Catalot, V., C. Laredo, M. Nussbaum (2002). 'Asymptotic equivalence of estimating a Poisson intensity and a positive diffusion drift'. In: *Ann. Statist.* 30.3. Dedicated to the memory of Lucien Le Cam, pp. 731–753.

Gibbs, A. L., F. E. Su (2002). 'On choosing and bounding probability metrics'. In: *International statistical review* 70.3, pp. 419–435.

Golubev, G. K., M. Nussbaum, H. H. Zhou (2010). 'Asymptotic equivalence of spectral density estimation and Gaussian white noise'. In: *Ann. Statist.* 38.1, pp. 181–214.

Grama, I., M. Nussbaum (1998). 'Asymptotic equivalence for nonparametric generalized linear models'. In: *Probab. Theory Related Fields* 111.2, pp. 167–214.

— (2002). 'Asymptotic equivalence for nonparametric regression'. In: *Math. Methods Statist.* 11.1, pp. 1–36.

Grama, I. G., M. H. Neumann (2006). 'Asymptotic equivalence of nonparametric autoregression and nonparametric regression'. In: *Ann. Statist.* 34.4, pp. 1701–1732.

Guy, R. (2013). 'Inférence dans le cadre des maladies transmissibles par des processus de diffusion'. In: *Ph.D thesis*.

Hansen, O. H., E. N. Torgersen (1974). 'Comparison of linear normal experiments'. In: *The Annals of Statistics*, pp. 367–373.

Hellinger, E. (1907). *Die Orthogonalinvarianten quadratischer Formen von unendlichvielen Variabelen*. W. Fr. Kaestner.

Hoffmann, M. (1999). 'Adaptive estimation in diffusion processes'. In: *Stochastic processes and their Applications* 79.1, pp. 135–163.

Ibragimov, I., R. Khas' Minskii (1991). 'Asymptotically normal families of distributions and efficient estimation'. In: *The Annals of Statistics*, pp. 1681–1724.

Ibragimov, I. A., R. Z. Has'minskii (1981). *Statistical estimation*. Springer.

Jacod, J., A. N. Shiryaev (1987). *Limit theorems for stochastic processes*. Vol. 288. Springer-Verlag Berlin.

Jacod, J., A. N. Shiryaev (2003). *Limit theorems for stochastic processes*. Second. Vol. 288. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, pp. xx+661.

Jähnisch, M., M. Nussbaum (2003). 'Asymptotic equivalence for a model of independent non identically distributed observations'. In: *Statist. Decisions* 21.3, pp. 197–218.

Kakutani, S. (1948). 'On equivalence of infinite product measures'. In: *Annals of Mathematics*, pp. 214–224.

Karatzas, I., S. E. Shreve (2000). *Brownian motion and stochastic calculus*. Second. Vol. 113. Graduate Texts in Mathematics. Springer-Verlag, New York, pp. xxiv+470.

Kullback, S., R. A. Leibler (1951). 'On information and sufficiency'. In: *The Annals of Mathematical Statistics*, pp. 79–86.

Kutoyants, Y. A. (1984a). 'On nonparametric estimation of trend coefficients in a diffusion process'. In: *Statistics and Control of Stochastic Processes* 230, p. 250.

— (1984b). *Parameter estimation for stochastic processes*. Vol. 6. Heldermann.

Laredo, C. F. (1990). 'A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process'. In: *The Annals of Statistics*, pp. 1158–1171.

Le Cam, L. (1964). 'Sufficiency and approximate sufficiency'. In: *Ann. Math. Statist.* 35, pp. 1419–1455.

Le Cam, L. (1969). *Théorie asymptotique de la décision statistique*. Séminaire de Mathématiques Supérieures, No 33 (Été, 1968). Les Presses de l'Université de Montréal, Montreal, Que., p. 140.

— (1986). *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. New York: Springer-Verlag, pp. xxvi+742.

Le Cam, L., G. L. Yang (2000). *Asymptotics in statistics*. Second. Springer Series in Statistics. Some basic concepts. New York: Springer-Verlag, pp. xiv+285.

Liese, F. (1987). 'Estimates of Hellinger integrals of infinitely divisible distributions'. In: *Kybernetika* 23.3, pp. 227–238.

Liese, F., I. Vajda (1987). *Convex statistical distances*. Vol. 95. Teubner-Texte zur Mathematik [Teubner Texts in Mathematics]. With German, French and Russian summaries. BSB B. G. Teubner Verlagsgesellschaft, Leipzig, p. 224.

Mariucci, E. (2015a). 'Asymptotic equivalence for density estimation and Gaussian white noise: an extension'. In: *arXiv preprint arXiv:1503.05019*.

— (2015b). 'Asymptotic equivalence for inhomogeneous jump diffusion processes and white noise'. In: *ESAIM: Probability and Statistics*. DOI:10.1007/s11203-015-9117-x.

— (2015c). 'Asymptotic equivalence of discretely observed diffusion processes and their Euler scheme: small variance case'. In: *Stat. Inference Stoch. Process.* DOI:10.1007/s11203-015-9117-x.

— (2015d). 'Asymptotic equivalence of Lévy density estimation and Gaussian white noise'. In: *arXiv preprint arXiv:1503.04530*.

Meister, A. (2011). 'Asymptotic equivalence of functional linear regression and a white noise inverse problem'. In: *Ann. Statist.* 39.3, pp. 1471–1495.

Meister, A., M. Reiß (2013). 'Asymptotic equivalence for nonparametric regression with non-regular errors'. In: *Probab. Theory Related Fields* 155.1-2, pp. 201–229.

Memin, J, A. Shiryayev (1985). 'Distance de Hellinger-Kakutani des lois correspondant à deux processus à accroissements indépendants'. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 70.1, pp. 67–89.

Milstein, G., M. Nussbaum (1998). 'Diffusion approximation for nonparametric autoregression'. In: *Probab. Theory Related Fields* 112.4, pp. 535–543.

Newman, C. M. (1972). 'The inner product of path space measures corresponding to random processes with independent increments'. In: *Bulletin of the American Mathematical Society* 78.2, pp. 268–271.

Nourdin, I., G. Peccati (2009). 'Stein's method on Wiener chaos'. In: *Probability Theory and Related Fields* 145.1-2, pp. 75–118.

Nussbaum, M. (1996). 'Asymptotic equivalence of density estimation and Gaussian white noise'. In: *Ann. Statist.* 24.6, pp. 2399–2430.

Øksendal, B. (1985). *Stochastic differential equations*. Universitext. An introduction with applications. Berlin: Springer-Verlag, pp. xiii+205.

Peccati, G. (2011). 'The Chen-Stein method for Poisson functionals'. In: *arXiv preprint arXiv:1112.5051*.

Picard, J. (1986). 'Nonlinear filtering of one-dimensional diffusions in the case of a high signal-to-noise ratio'. In: *SIAM Journal on Applied Mathematics* 46.6, pp. 1098–1125.

— (1991). 'Efficiency of the extended Kalman filter for nonlinear systems with small noise'. In: *SIAM Journal on Applied Mathematics* 51.3, pp. 843–885.

Reiß, M. (2008). 'Asymptotic equivalence for nonparametric regression with multivariate and random design'. In: *Ann. Statist.* 36.4, pp. 1957–1982.

— (2011). 'Asymptotic equivalence for inference on the volatility from noisy observations'. In: *Ann. Statist.* 39.2, pp. 772–802.

Reiss, R.-D. (1989). *Approximate distributions of order statistics*. Springer.

Revuz, D., M. Yor (1999). *Continuous martingales and Brownian motion*. Vol. 293. Springer Science & Business Media.

Reynaud-Bouret, P. (2003). 'Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities'. In: *Probability Theory and Related Fields* 126.1, pp. 103–153.

Rohde, A. (2004). 'On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise'. In: *Statist. Decisions* 22.3, pp. 235–243.

Ross, N. (2011). 'Fundamentals of Stein's method'. In: *Probab. Surv* 8, pp. 210–293.

Sato, K.-i. (1999). *Lévy processes and infinitely divisible distributions*. Vol. 68. Cambridge Studies in Advanced Mathematics. Translated from the 1990 Japanese original, Revised by the author. Cambridge: Cambridge University Press, pp. xii+486.

— (2000). 'Density transformation in Lévy processes'. In:

Schmidt-Hieber, J. (2014). 'Asymptotic equivalence for regression under fractional noise'. In: *The Annals of Statistics* 42.6, pp. 2557–2585.

Shannon, C. E. (1948). 'A mathematical theory of communication'. In: *Bel System Tech. J.* 27, pp. 379–423.

Shiryaev, A. N., V. G. Spokoiny (2000a). *Statistical experiments and decisions*. Vol. 8. Advanced Series on Statistical Science & Applied Probability. Asymptotic theory. River Edge, NJ: World Scientific Publishing Co. Inc., pp. xvi+281.

— (2000b). *Statistical experiments and decisions*. Vol. 8. Advanced Series on Statistical Science & Applied Probability. Asymptotic theory. River Edge, NJ: World Scientific Publishing Co. Inc., pp. xvi+281.

Strasser, H. (1985). *Mathematical theory of statistics*. Vol. 7. de Gruyter Studies in Mathematics. Statistical experiments and asymptotic decision theory. Berlin: Walter de Gruyter & Co., pp. xii+492.

Strauch, C. (2015). 'Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions'. In: *Probability Theory and Related Fields*, pp. 1–40.

Torgersen, E. N. (1972). 'Comparison of translation experiments'. In: *The Annals of Mathematical Statistics* 43.5, pp. 1383–1399.

— (1974). 'Comparison of experiments by factorization'. In: *Stat.Res.Report, Univ. of Oslo*.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Springer Series in Statistics. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. Springer, New York, pp. xii+214.

Uchida, M., N. Yoshida (2004). 'Asymptotic expansion for small diffusions applied to option pricing'. In: *Stat. Inference Stoch. Process.* 7.3, pp. 189–223.

Volkonskii, V. (1958). 'Random substitution of time in strong Markov processes'. In: *Theory of Probability & Its Applications* 3.3, pp. 310–326.

Wald, A. (1939). 'Contributions to the theory of statistical estimation and testing hypotheses'. In: *Ann. Math. Statistics* 10, pp. 299–326.

— (1950). 'Statistical decision functions.' In:

Wang, Y. (2002a). 'Asymptotic nonequivalence of Garch models and diffusions'. In: *Ann. Statist.* 30.3. Dedicated to the memory of Lucien Le Cam, pp. 754–783.

— (2002b). 'Asymptotic nonequivalence of Garch models and diffusions'. In: *Ann. Statist.* 30.3. Dedicated to the memory of Lucien Le Cam, pp. 754–783.

Zolotarev, V. M. (1983). 'Probability metrics'. In: *Teoriya Veroyatnostei i ee Primeneniya* 28.2, pp. 264–287.